

UNIVERSITÉ DENIS DIDEROT PARIS VII

THÈSE

POUR OBTENIR LE GRADE DE

DOCTEUR DE L'UNIVERSITÉ PARIS VII

Spécialité : Mathématiques Appliquées

présentée par

Claire CANNAMÉLA

---

APPORT DES MÉTHODES PROBABILISTES DANS  
LA SIMULATION DU COMPORTEMENT SOUS IRRADIATION  
DU COMBUSTIBLE À PARTICULES

---

Soutenue le 21 Septembre 2007, devant le jury composé de :

PRÉSIDENT	M. Jean-Claude FORT	Université Toulouse III
DIRECTEUR DE THÈSE	M. Josselin GARNIER	Université Paris VII
RAPPORTEURS	M. Bernard BERCU	Université de Bordeaux
	M. Pierre DEL MORAL	Université de Nice
EXAMINATEURS	M. Nicolas DEVICTOR	CEA Cadarache
	M. Frédéric MICHEL	CEA Cadarache

Thèse préparée au Commissariat à l'Énergie Atomique de Cadarache.



## REMERCIEMENTS

Le moment est venu d'exprimer ma reconnaissance à toutes les personnes, qui de près ou de loin, ont contribué à mener à bien ce travail de thèse.

Je tiens à remercier en premier lieu mon directeur de thèse Josselin GARNIER tant pour ses compétences scientifiques que pour ses qualités humaines. Malgré son éloignement géographique, sa disponibilité, ses conseils avisés et sa patience ont largement contribué à l'aboutissement de ce travail.

Je souhaite remercier les membres du jury qui ont consacré une partie de leur temps pour s'intéresser à mon travail de recherche. Je suis honorée que Pierre DEL MORAL et Bernard BERCU aient accepté d'être rapporteurs de ma thèse et je les remercie pour leur lecture attentive du manuscrit. Je tiens également à remercier Jean-Claude FORT qui m'a fait l'honneur de présider mon jury de thèse.

J'adresse mes remerciements à mes deux encadrants CEA. À Michel PELLETIER pour ses discussions au cours desquelles il m'a fait partager sa vision claire et synthétique des enjeux du nucléaire. À Frédéric MICHEL pour s'être intéressé de près aux différentes méthodes probabilistes abordées dans cette thèse et pour son efficace aide finale.

J'exprime mes plus sincères remerciements à Nicolas DEVICTOR qui a toujours fait preuve d'une grande disponibilité, et ce malgré ses nombreuses occupations. Ses conseils et son soutien m'ont été d'une précieuse aide. Un grand merci aussi d'avoir accepté de faire parti de mon jury de thèse.

Je suis reconnaissante à Laurence BRUNEL et Patrick OBRY, respectivement chef du service d'étude et de simulation du comportement des combustibles et chef du laboratoire de simulation du comportement des combustibles, pour m'avoir fait confiance en m'accueillant au sein de leur équipe.

Un grand merci aussi à Bertrand IOOSS pour sa gentillesse et sa générosité. Il a toujours su m'accorder du temps quand j'en avais besoin et m'a toujours soutenue.

Merci à tous les membres du LSC pour leur bonne humeur. Je remercie particulièrement Etienne CASTELIER qui a rendu mon environnement informatique convivial. Un grand merci à Régine BOUSQUET pour sa gentillesse.

Je remercie également les personnes qui travaillent sur le combustible à particules et les

HTR, ils ont tous contribué à la compréhension de l'application nucléaire : Mayeul PHELIP (chargé d'affaire au moment de la thèse), Pierre GUILLERMIER (AREVA NP),...

Je souhaite aussi saluer toutes les personnes avec qui j'ai partagé le bureau : Thomas et Sophie au début qui m'ont donné des conseils de thésards avertis... A Gilles, stagiaire très doué avec qui j'ai travaillé sur l'analyse de sensibilité. Et bon courage à Victor et Jérôme pour la fin de leur thèse.

Aux membres du LLCC et à J:m, merci pour les repas de midi et les pauses café qui se sont révélées très sympathiques et divertissantes.

L'occasion m'est donnée ici de remercier tous ceux qui m'ont permis de m'évader. D'abord avec ce superbe voyage au Kirghizstan : Ln, Yannou et Fred, merci pour ces fabuleuses semaines, une pensée particulière à Manue dont c'était l'initiative mais qui n'a malheureusement pas pu se joindre à nous. Ensuite à cheval, ces quelques randonnées au bord du Verdon sont de très bons souvenirs. Sur la neige aussi, grâce au ski de randonnée, dur mais si bon finalement, merci à JiB, Fanfou, Bambi, et à tous les passionnés de montagne de Cadarache. Et puis dans l'eau, merci aux membres du club de plongée grâce à qui j'ai découvert un monde subaquatique merveilleux et plein de surprises. Et particulièrement à Laure pour nos fous rires plein de bulles... Et aussi à Célia et Anne-So pour leur amitié et nos nombreuses soirées filles, à Jo, toujours très attentive aux autres ; sans oublier Cyril toujours partant et Valérie qui a pris soin de moi. Merci pour votre soutien et vos témoignages d'affection. Merci également à mes amis de longue date (10 ans déjà !), Clément et Magali, éparpillés malheureusement mais qui m'ont soutenue à distance.

Merci à mes parents et parents par alliance qui m'ont stimulée et encouragée pendant mes études. Ils m'ont toujours donné le choix et m'ont supportée dans mes décisions. Une pensée pour mes deux sœurs : bon courage à May pour ses nouvelles responsabilités et bonne chance à Lise pour la vie qui s'ouvre devant elle...

Merci à LGC d'avoir été si présent sur la fin de la thèse, de m'avoir fait confiance, soutenue, motivée parfois, écoutée, encouragée. Je suis heureuse d'avoir partagé ces moments avec toi, et tous les autres aussi.

*À ma famille,  
je dédie cette thèse.*

## TABLE DES MATIÈRES

<b>Introduction</b>	<b>1</b>
<b>Chapitre 1 : Méthodes num. d'intégration en fiabilité des structures</b>	<b>5</b>
1.1 Traitement des incertitudes . . . . .	6
1.1.1 Modèle et incertitudes . . . . .	6
1.1.1.1 Approche inférentielle classique . . . . .	6
1.1.1.2 Approche bayésienne . . . . .	7
1.1.2 Analyse de sensibilité . . . . .	7
1.2 Indices de fiabilité et méthodes d'approximation . . . . .	8
1.2.1 Indices de fiabilité . . . . .	8
1.2.2 Méthodes FORM/SORM . . . . .	11
1.2.2.1 Changement de l'espace d'entrée des variables . . . . .	11
1.2.2.2 Recherche des points de conception . . . . .	12
1.2.2.3 Approximation de la surface d'état limite . . . . .	14
1.2.2.4 Calcul de la probabilité de défaillance . . . . .	16
1.2.2.5 Fiabilité des systèmes . . . . .	17
1.2.2.6 Quelques remarques . . . . .	18
1.2.3 Couplage mécano-fiabiliste . . . . .	19
1.3 Méthodes de simulation . . . . .	22
1.3.1 Monte Carlo direct . . . . .	22
1.3.1.1 Contexte général . . . . .	22
1.3.1.2 Fiabilité des structures . . . . .	23
1.3.2 Monte Carlo conditionnel . . . . .	25
1.3.2.1 Contexte général . . . . .	25
1.3.2.2 Contexte fiabilité des structures . . . . .	26
1.3.3 Corrélation de variables . . . . .	27
1.3.3.1 Variables corrélées . . . . .	27
1.3.3.2 Variables antithétiques . . . . .	28

1.3.3.3	Variables de contrôle . . . . .	28
1.3.4	Stratification . . . . .	29
1.3.4.1	Contexte général . . . . .	29
1.3.4.2	Contexte fiabilité des structures . . . . .	30
1.3.4.3	Hypercube latin . . . . .	30
1.3.5	Simulation par somme de Riemann . . . . .	32
1.3.6	Tirage d'importance . . . . .	32
1.3.6.1	Contexte général . . . . .	32
1.3.6.2	Contexte fiabilité des structures . . . . .	33
1.3.6.3	Découpage du domaine d'intégration . . . . .	34
1.3.7	Simulation directionnelle . . . . .	35
1.3.8	Méthodes de quasi Monte Carlo . . . . .	36
1.3.9	Combinaisons de méthodes . . . . .	38
1.3.9.1	Avec la méthode de Monte Carlo conditionnel . . . . .	38
1.3.9.2	Méthodes d'approximations et de simulations . . . . .	38
1.3.9.3	Tirage directionnel et tirage d'importance . . . . .	38
1.3.9.4	Simulation axis orthogonale . . . . .	39
1.3.10	Organigramme des méthodes . . . . .	40
1.4	Synthèse partielle . . . . .	40
 <b>Chapitre 2 : Tirage d'importance</b>		<b>43</b>
2.1	Généralités sur le tirage d'importance . . . . .	43
2.1.1	Résultats classiques . . . . .	43
2.1.2	Variantes de l'estimateur . . . . .	45
2.1.2.1	Première variante . . . . .	45
2.1.2.2	Deuxième variante . . . . .	46
2.1.3	Autres utilisations du tirage d'importance . . . . .	46
2.2	État de l'art . . . . .	47
2.2.1	Tirage d'importance pour la fiabilité des structures . . . . .	47
2.2.1.1	Densité d'importance centrée autour du point de conception . . . . .	47
2.2.1.2	Tirage d'importance conditionné . . . . .	48
2.2.1.3	Tirage d'importance asymptotique . . . . .	48
2.2.1.4	Mise à jour de la probabilité approchée . . . . .	49

2.2.1.5	Approche adaptative de Bucher . . . . .	50
2.2.1.6	Approche adaptative de Karamchandani, Bjerager et Cornell . . . . .	51
2.2.1.7	Technique adaptative de Wu . . . . .	52
2.2.1.8	Méthodes utilisant des noyaux . . . . .	53
2.2.1.9	Tirage d'importance filtré avec SVM . . . . .	54
2.2.1.10	Tirage d'importance pour la fiabilité des systèmes . . . . .	55
2.2.1.11	Cas fiabilité des systèmes et calcul FORM . . . . .	56
2.2.1.12	Comparaison de l'efficacité des méthodes . . . . .	58
2.2.2	Construction de quelques densités d'importance dans un cadre général . . . . .	59
2.2.2.1	Le rapport de vraisemblance . . . . .	59
2.2.2.2	Méthodes non paramétriques . . . . .	60
2.2.2.3	Approche par mélange . . . . .	61
2.3	Tirage d'importance adaptatif . . . . .	62
2.3.1	Construction d'un estimateur adaptatif . . . . .	63
2.3.1.1	Estimateur simple à k étapes . . . . .	63
2.3.1.2	Estimateur avec recyclage à k étapes . . . . .	64
2.3.1.3	Estimateur mélange . . . . .	68
2.3.2	Recherche de la densité d'importance . . . . .	68
2.3.2.1	Optimisation sur les moments de la densité paramétrique . . . . .	69
2.3.2.2	Minimisation de la variance . . . . .	71
2.3.2.3	Minimisation de l'entropie croisée . . . . .	74
2.3.3	Critère d'optimisation de $n_k$ . . . . .	75
2.3.3.1	Comportement asymptotique de $\lambda_1$ . . . . .	75
2.3.3.2	Choix de $n_1$ et $n_2$ . . . . .	77
2.3.3.3	Illustrations . . . . .	79
2.4	Simulations sur des cas tests . . . . .	85
2.4.1	Fonctions simples utilisées pour tester les méthodes . . . . .	85
2.4.2	Recherche de la densité d'importance . . . . .	86
2.4.2.1	Estimateur empirique des moments avec recyclage . . . . .	87
2.4.2.2	Augmentation de la variance . . . . .	89
2.4.2.3	Etapes intermédiaires sans recyclage . . . . .	90
2.4.2.4	Efficacité des estimateurs . . . . .	91
2.4.3	Performance et optimisation de $\hat{I}$ pour un nombre limité de tirages . . . . .	92

2.4.3.1	Test du nombre d'étapes . . . . .	92
2.4.3.2	limitation des poids . . . . .	93
2.4.3.3	Utilisation du bootstrap sur un estimateur empirique . . . . .	95
2.4.3.4	Recyclage à partir d'une densité mixte . . . . .	95
2.4.4	Conclusion sur les estimateurs . . . . .	96
<b>Chapitre 3 : Simulation multi-niveaux</b>		<b>97</b>
3.1	L'idée de base de la simulation multi-niveaux . . . . .	97
3.2	les méthodes MCMC . . . . .	98
3.2.1	Algorithme de Metropolis-Hasting . . . . .	98
3.2.2	Gibbs . . . . .	99
3.2.3	Convergence des méthodes MCMC . . . . .	99
3.2.4	Algorithme pour simuler $f(\cdot, \mathbf{D}_i)$ , $i = 1, \dots, m - 1$ . . . . .	99
3.3	Estimateur des probabilités conditionnelles . . . . .	100
3.3.1	Estimateur prenant en compte tous les états des chaînes . . . . .	100
3.3.2	Estimateur prenant en compte quelques états des chaînes . . . . .	101
3.3.2.1	Aperçu des méthodes de splitting . . . . .	101
3.3.2.2	Estimateur par une méthode de splitting . . . . .	102
3.4	Choix des paramètres importants . . . . .	103
3.4.1	les niveaux intermédiaires . . . . .	103
3.4.2	La loi instrumentale . . . . .	104
3.5	Forme particulière du domaine de défaillance . . . . .	105
3.5.1	Isolement d'une variable . . . . .	105
3.5.2	Propriétés de l'estimateur . . . . .	106
3.6	Simulations sur des cas tests . . . . .	107
3.6.1	Exemple 1 . . . . .	107
3.6.2	Exemple 2 . . . . .	108
3.6.3	Exemple 3 . . . . .	109
3.6.4	Exemple 4 . . . . .	109
3.7	Conclusion . . . . .	111
<b>Chapitre 4 : Application au combustible HTR</b>		<b>113</b>
4.1	Contexte industriel . . . . .	113
4.1.1	Le réacteur à haute température . . . . .	113

4.1.2	Le combustible à particules . . . . .	115
4.1.2.1	Présentation du combustible . . . . .	115
4.1.2.2	Comportement sous irradiation . . . . .	117
4.1.2.3	Les modes de ruptures de la particule . . . . .	119
4.2	Codes modélisant le comportement d'une particule . . . . .	120
4.2.1	Le code ATLAS . . . . .	120
4.2.1.1	Modélisation de comportement d'une particule . . . . .	120
4.2.1.2	Historique d'irradiation . . . . .	121
4.2.1.3	Calcul de la probabilité de rupture . . . . .	122
4.2.2	Les codes internationaux . . . . .	124
4.2.2.1	Modélisations basées sur une seule particule . . . . .	125
4.2.2.2	Modèles analytiques utilisant des méthodes statistiques . . . . .	126
4.2.2.3	Utilisation des éléments finis dans la modélisation . . . . .	128
4.3	Analyse de sensibilité du code ATLAS . . . . .	129
4.3.1	Incertitude en entrée . . . . .	130
4.3.1.1	Les paramètres de fabrication . . . . .	130
4.3.1.2	Les paramètres d'irradiation . . . . .	130
4.3.1.3	Lois de comportement . . . . .	131
4.3.2	Caractéristiques des réacteurs et expériences étudiés . . . . .	132
4.3.3	Résultats . . . . .	133
4.3.3.1	HFR-EU2 . . . . .	134
4.3.3.2	GT-MHR . . . . .	136
4.4	Fiabilité du combustible . . . . .	138
4.4.1	HFR-EU1 . . . . .	138
4.4.2	PBMR . . . . .	141
4.4.2.1	Gestion du combustible . . . . .	141
4.4.2.2	Conditions en réacteur . . . . .	142
4.4.2.3	Résultats . . . . .	143
4.4.3	GT-MHR . . . . .	145
4.4.3.1	Gestion du combustible . . . . .	146
4.4.3.2	Conditions en réacteur . . . . .	146
4.4.3.3	Résultats . . . . .	148

<b>Chapitre 5 :</b>	<b>Estimation de quantiles et application</b>	<b>151</b>
5.1	Introduction . . . . .	151
5.2	Estimation de quantile sur un $n$ -échantillon . . . . .	152
5.2.1	Quantile empirique . . . . .	153
5.2.2	Quantile de Wilks . . . . .	154
5.2.3	Quantile empirique avec bootstrap . . . . .	156
5.2.4	Autres résultats sur les dépassements de seuil . . . . .	159
5.2.4.1	Prédiction de dépassement d'un seuil . . . . .	159
5.2.4.2	Temps d'attente d'un dépassement . . . . .	159
5.3	Estimation de quantile par variable de contrôle . . . . .	160
5.3.1	Estimation de la fonction de répartition par max. de vraisemblance . . .	160
5.3.2	Estimation de la fonction de répartition avec contrôle . . . . .	161
5.3.3	Estimation de quantile . . . . .	163
5.3.4	L'estimateur avec contrôle optimal . . . . .	163
5.4	Estimation de quantile par stratification contrôlée . . . . .	165
5.4.1	Estimation de la fonction de répartition . . . . .	165
5.4.1.1	Estimateur avec stratification contrôlée . . . . .	165
5.4.1.2	Estimation de la fonction de répartition . . . . .	166
5.4.1.3	Estimation de la queue de la fonction de répartition . . . . .	167
5.4.2	Estimation d'un quantile . . . . .	168
5.4.2.1	Mise en oeuvre . . . . .	168
5.4.2.2	Simulations . . . . .	168
5.4.3	Stratification contrôlée adaptative . . . . .	173
5.4.3.1	Choix optimal du nombre de simulations par strates . . . . .	173
5.4.3.2	Estimateur adaptatif de la fonction de répartition . . . . .	175
5.4.3.3	Réduction de variance . . . . .	175
5.4.4	Stratification contrôlée adaptative pour l'estimation d'un quantile . . .	177
5.4.4.1	Mise en oeuvre . . . . .	177
5.4.4.2	Simulations . . . . .	179
5.5	Estimation de quantile avec échantillonnage préférentiel . . . . .	181
5.5.1	Estimation de la fonction de répartition . . . . .	181
5.5.1.1	Estimation avec échantillonnage préférentiel . . . . .	181
5.5.1.2	Recherche de la densité d'importance . . . . .	182

5.5.2	Estimation du quantile à partir du modèle réduit . . . . .	183
5.5.3	Simulations . . . . .	183
5.6	Implémentation avec le code CATHARE . . . . .	185
5.6.1	Présentation du code . . . . .	185
5.6.2	Détermination d'une surface de réponse . . . . .	186
5.6.2.1	Screening . . . . .	186
5.6.2.2	Régression . . . . .	187
5.6.2.3	Analyse de la validité de la surface de réponse . . . . .	189
5.6.3	Stratification contrôlée . . . . .	191
5.6.3.1	Stratégie à quatre strates . . . . .	191
5.6.3.2	Stratégie à deux strates . . . . .	192
5.6.3.3	Stratégie à cinq strates . . . . .	193
5.6.4	Echantillonnage préférentiel . . . . .	194
5.6.5	Comparaisons à un estimateur empirique . . . . .	195
5.7	Conclusion . . . . .	197
	<b>Conclusion</b>	<b>201</b>
	<b>Références</b>	<b>205</b>
	<b>Annexe A : Méthodes d'analyse de sensibilité</b>	<b>219</b>
A.1	Étude d'un modèle linéaire . . . . .	219
A.2	Méthode de « screening » de MORRIS . . . . .	221
A.3	Méthode de SOBOL . . . . .	224
A.4	Méthode FAST . . . . .	227
A.5	Conclusion : dans la pratique . . . . .	228

## LISTE DES FIGURES

1.2.1 Illustration du problème $R - S$ . . . . .	9
1.2.2 Transformation de l'espace d'entrée des variables . . . . .	12
1.2.3 Méthodes FORM et SORM . . . . .	16
1.2.4 Principe du couplage direct . . . . .	19
1.3.1 Monte Carlo direct . . . . .	24
1.3.2 Monte Carlo conditionnel . . . . .	27
1.3.3 Monte Carlo stratifié . . . . .	31
1.3.4 Monte Carlo Hypercube Latin . . . . .	31
1.3.5 Monte Carlo Tirage d'importance . . . . .	34
1.3.6 Simulation directionnelle . . . . .	36
1.3.7 Organigramme des méthodes . . . . .	40
2.2.1 Tirage d'importance conditionné . . . . .	49
2.3.1 Évolution de $\Theta^2$ en fonction des paramètres initiaux . . . . .	80
2.3.2 lignes de niveau de $\Theta^2$ en fonction et $n_1^c$ en fonction des paramètres initiaux . . . . .	81
2.3.3 Évolution de $\mathbb{E}[\Omega^2]/n_2$ en fonction de $n_1$ - cas 1 . . . . .	82
2.3.4 Évolution de $\mathbb{E}[\Omega^2]/n_2$ en fonction de $n_1$ - cas 2 . . . . .	83
2.3.5 Évolution de $\mathbb{E}[\Omega^2]/n_2$ en fonction de $n_1$ - cas 3 . . . . .	84
2.4.1 Fonction d'importance optimale $f^*$ pour les exemples 1, 2 et 3 . . . . .	86
2.4.2 Graphe des poids à l'étape 1 - exemple 2 . . . . .	88
2.4.3 Évolution des fonctions d'importances - exemple 2 et 3 . . . . .	88
2.4.4 Illustration du cas d'un poids élevé à l'issue de l'étape 1 . . . . .	89
2.4.5 Illustration du cas d'un poids élevé à l'issue de l'étape 2, mauvaise direction . . . . .	89
2.4.6 Recherche de la densité d'importance, variance doublée - exemple 2 . . . . .	90
2.4.7 Résultats de l'estimation finale, variance doublée, sans recyclage - exemple 2 . . . . .	91
2.4.8 Résultats de l'estimation finale, variance doublée, sans recyclage - exemple 3 . . . . .	92
2.4.9 Influence d'une mauvaise estimation de la moyenne sur l'estimation finale . . . . .	92
2.4.10 Suites de densités d'importance, mauvaise estimation finale - exemple 2 . . . . .	92
2.4.11 Histogramme des moyennes des densités d'importance - exemple 2 . . . . .	94

3.2.1	Illustration de l’algorithme de Metropolis Hasting modifié . . . . .	101
3.3.1	Illustration de la méthode de splitting en dimension 1 . . . . .	102
3.3.2	Modèle de branchement avec duplication de trajectoires en dimension 1 . . . . .	103
3.4.1	Choix adaptatifs des domaines intermédiaires . . . . .	104
3.6.1	Évolution des chaînes de Markov, exemple 1d . . . . .	108
3.6.2	Histogramme de probabilités de rupture, exemple 1d . . . . .	108
3.6.3	Évolution des chaînes de Markov, exemple 2d . . . . .	109
3.6.4	Histogramme de probabilités de rupture, exemple 2d . . . . .	109
3.6.5	Histogramme de probabilités de rupture, exemple 4d . . . . .	110
3.6.6	Comparaison des estimateurs . . . . .	110
4.1.1	Réacteur VHTR . . . . .	114
4.1.2	Le combustible à particules . . . . .	115
4.1.3	Les assemblages combustible . . . . .	116
4.1.4	Comportement sous irradiation de la particule . . . . .	118
4.1.5	Évolution des contraintes au cours du temps . . . . .	118
4.1.6	Les modes de la rupture de particules . . . . .	119
4.2.1	Géométrie 1D du code ATLAS . . . . .	121
4.2.2	Géométrie 2D et 3D du code ATLAS . . . . .	122
4.2.3	Exemple d’historique d’irradiation . . . . .	122
4.2.4	code PANAMA - contrainte dans le SiC et probabilité de rupture . . . . .	125
4.3.1	HFR-EU2 - Analyse d’incertitude . . . . .	134
4.3.2	GT-MHR - Analyse d’incertitude . . . . .	136
4.3.3	Réacteur GT-MHR : Analyse de sensibilité . . . . .	138
4.4.1	HFR-EU1 - Résultats du code ATLAS . . . . .	139
4.4.2	HFR-EU1 : densités d’importance de l’épaisseur des couches . . . . .	140
4.4.3	HFR-EU1 : probabilités de rupture (polynôme) . . . . .	141
4.4.4	Temps de passage des boulets en fonction de leur place dans le coeur . . . . .	142
4.4.5	Facteur correctif de la vitesse de fission et de la fluence rapide . . . . .	142
4.4.6	PBMR - Résultats ATLAS (contraintes) . . . . .	144
4.4.7	Histogramme de la probabilité de rupture estimée par TI - PBMR (essai 1) . . . . .	145
4.4.8	Histogramme de la probabilité de rupture estimée par TI - PBMR (essai 1) . . . . .	146
4.4.9	GT-MHR - Résultats ATLAS . . . . .	149

4.4.1	Histogramme de la probabilité de rupture estimée par TI - GT-MHR . . . . .	150
5.2.1	Comparaisons entre l'estimateur empirique et l'estimateur de Wilks . . . . .	157
5.4.1	Comparaison du quantile empirique et quantile avec SC - fonction 1D . . . . .	169
5.4.2	Comparaison du quantile empirique et quantile avec SC - fonction 2D . . . . .	170
5.4.3	Comparaison du quantile empirique et quantile avec SC - fonction d'Ishigami .	172
5.4.4	Comparaison du quantile empirique et quantile avec SC - fonction de Morris . .	173
5.5.1	Comparaison du quantile empirique et quantile avec EP - fonction 1D . . . . .	184
5.5.2	Comparaison du quantile empirique et quantile avec EP - fonction 2D . . . . .	185
5.6.1	Comparaison du code et de la surface de réponse . . . . .	190
5.6.2	Histogramme du quantile estimé avec SC (4 strates) - code CATHARE . . . . .	192
5.6.3	Histogramme du quantile estimé avec SC (2 strates) - code CATHARE . . . . .	193
5.6.4	Histogramme du quantile estimé avec SC (5 strates) - code CATHARE . . . . .	194
5.6.5	Histogramme du quantile estimé avec EP - code CATHARE . . . . .	196

## LISTE DES TABLEAUX

1.3.1 Vitesses de convergence en dimension $d$ pour des fonctions régulières . . . . .	23
2.4.1 Comparaison des estimateurs, tirage d'importance - exemple 3 . . . . .	93
2.4.2 Comparaison des estimateurs, tirage d'importance (bootstrap) - exemple 3 . . . . .	95
2.4.3 Estimateurs des paramètres de la densité d'importance - exemple 3 . . . . .	96
4.2.1 Quelques uns des codes internationaux . . . . .	124
4.3.1 Paramètres de fabrication des particules . . . . .	130
4.3.2 Lois de probabilité des paramètres de fabrication . . . . .	131
4.3.3 Paramètres d'irradiation . . . . .	131
4.3.4 Lois de comportement . . . . .	132
4.3.5 Conditions d'irradiation des réacteurs HFR et GTM-HR . . . . .	133
4.3.6 Lois de probabilité des lois de comportement . . . . .	133
4.3.7 Réacteur HFR-EU2 : Analyse de sensibilité . . . . .	135
4.3.8 Réacteur GT-MHR : Analyse de sensibilité . . . . .	137
4.4.1 HFR-EU1 : densités initiales et d'importance . . . . .	140
4.4.2 HFR-EU1 : comparaison du coefficient de variation . . . . .	141
4.4.3 PBMR : Paramètres aléatoires . . . . .	144
4.4.4 PBMR - Probabilité de rupture (essai 1) . . . . .	144
4.4.5 PBMR - Probabilité de rupture (essai 2) . . . . .	145
4.4.6 GT-MHR : Paramètres aléatoires . . . . .	148
4.4.7 GT-MHR - Probabilité de rupture . . . . .	149
5.6.1 Quantile empirique et écart-type en fonction de la taille d'échantillons . . . . .	197



## INTRODUCTION

### Contexte industriel

Nous recensons actuellement en France un parc de 58 réacteurs de deuxième génération, appartenant tous à la filière des réacteurs à eau pressurisée (REP). Ce parc doit être remplacé à partir de 2015 par des réacteurs de troisième génération (EPR), qui apporteront un gain en sûreté. Il est ensuite envisagé, mais pas avant 2030-2040, un passage à une quatrième génération qui devraient être plus économes, plus respectueux de l'environnement (minimisation des déchets), plus sûrs, plus fiables et s'ouvrant à d'autres applications industrielles (la production d'hydrogène, le dessalement de l'eau de mer, par exemple). Parmi les six concepts de quatrième génération se place le réacteur VHTR (Very High Temperature Reactor), réacteur modulaire à puissance modérée. Cette filière est plus adaptée à des pays en voie de développement dont le réseau électrique ne peut supporter des puissances unitaires élevées. Parmi les innovations majeures de ce type de réacteur, nous retrouvons l'emploi du combustible à particules (imaginé au début des années 50). Ce combustible se présente sous la forme de micro-billes enrobées dont le diamètre est de l'ordre du millimètre ; cette sphère est composée d'un noyau de matière fissile enrobé de différentes couches de matériaux céramiques réfractaires. Le CEA (DEC/SESC) a développé un logiciel déterministe, le code ATLAS, simulant le comportement thermique, physico-chimique et mécanique d'une particule sous irradiation. Il est incorporé à la plate-forme de simulation du combustible multi-filières PLEIADES. Une problématique particulièrement importante concerne la fiabilité des particules de combustible. En cas de rupture, un relâchement de produits de fission gazeux et volatiles peut se produire et contaminer le caloporteur du circuit primaire qui contaminerait à son tour le réacteur. Il est alors nécessaire de déterminer, à partir des résultats fournis par ATLAS la probabilité de rupture d'une population de particules en fonction du fonctionnement du réacteur, de la géométrie, des propriétés des matériaux, etc. C'est là l'objectif de cette thèse.

### La simulation

Grâce à l'essor, ces dernières années, des outils numériques, et les codes de calcul, devenus très performants, ont contribué au développement des méthodes de simulations. La simulation est la représentation du comportement d'un système réel à travers un modèle. Elle permet de

réaliser des expériences virtuelles difficilement réalisables dans le système réel, comme celle qui consiste à étudier le comportement dans des conditions extrêmes d'un système sans dégrader le système réel, ou encore celle qui étudie la manipulation d'échelle temporelle, c'est-à-dire l'étude « accélérée » d'un comportement. Nous nous intéressons ici au cas où un modèle désigne un code de calcul ; celui-ci renvoie, à partir de paramètres d'entrées, des réponses (ou sorties), ce sont les résultats des simulations (ou expériences simulées). La plupart de temps, les données d'entrées ne sont pas connues exactement, et ainsi une part d'incertitude réside en chacune d'elles, les sorties sont alors, elles aussi, entachées d'incertitude. Les simulations sont souvent gourmandes en ressources passives et en temps de calcul lorsque les modèles utilisés sont un tant soit peu « mécaniste » ou bien lorsque l'on souhaite obtenir des résultats avec un niveau de précision élevé. Les résultats de la simulation ne peuvent être qu'un échantillon d'un nombre potentiellement beaucoup plus grand d'observations. Il est donc d'un grand intérêt de mettre en place des techniques visant à accélérer le processus de simulation tout en préservant les nombreux avantages de cette méthode.

### **La fiabilité des structures**

Dans la présente étude, le domaine qui nous intéresse est celui de la fiabilité des structures. A l'origine, celle-ci concerne plutôt les ouvrages de génie civil de grandes tailles et les systèmes mécaniques de haute technologie (ponts, centrales nucléaires, barrages, ...).

Durant ces dernières années, l'étude mécano-fiabiliste est devenue un outil pour améliorer la qualité et surtout un outil d'aide à la décision (maintenance, réparations, ...). Un modèle de comportement d'une structure n'échappe pas à la présence d'incertitudes dues au modèle en lui-même mais aussi au manque de connaissance précise des lois de comportement des matériaux ou encore aux incertitudes de conception. La fiabilité des structures vise à étudier la fiabilité ou la défaillance d'un système en tenant compte de ces incertitudes par l'intermédiaire d'une estimation d'une probabilité de défaillance. Pour cela, elle s'appuie sur la définition de modes de défaillance, sur une modélisation des paramètres d'entrée (géométrie, matériau, chargement, ...) par des variables aléatoires, sur un modèle mécanique décrivant le comportement d'une structure et sur des outils fiabilistes tels que des indices de fiabilité, des méthodes d'approximations ou des méthodes de simulations.

Auparavant, la notion même de fiabilité reposait uniquement sur l'intuition, les compétences, l'expérience, et les retours d'expériences du ou des artisans. Cette démarche empirique a peu à peu laissé place à des démarches plus scientifiques. Dans la première moitié du XX<sup>e</sup> siècle est apparu *le principe des contraintes admissibles*, qui consistait à vérifier que les contraintes calculées en toute partie l'ouvrage et dans des conditions les plus défavorables,

restaient inférieures à une contrainte dite admissible. Cette dernière étant fixée sur la base des propriétés mécaniques des matériaux utilisés, elle s'obtenait en divisant la contrainte de rupture d'un matériau par un coefficient de sécurité. Cependant, la contrainte de rupture, significative s'il s'agit d'un matériau fragile, ne l'est plus dans le cas d'un matériau ductile (grande déformation inacceptable). C'est alors que vint la notion d'indice de fiabilité. Plusieurs personnes sont à l'origine de ce concept, Balser en 1961, Cornell en 1967, Lind en 1973 et Hasofer en 1974. De cette étape émergea *une démarche semi-probabiliste*, faisant intervenir, pour un état limite donné, une valeur caractéristique de l'effet des actions ou sollicitations  $S$ , une valeur caractéristique de la résistance  $R$  et des *coefficients partiels de sécurité* destinés à couvrir les incertitudes du modèle de la structure, des résistances, des sollicitations et des propriétés géométriques. Ensuite vint *la démarche probabiliste*, initiée par Madsen pour des problèmes d'ingénierie pétrolière, elle est une alternative aux approches semi-probabilistes et est maintenant bien connue sous l'appellation fiabilité des structures.

### Organisation de la thèse

Le premier chapitre présente un état de l'art des méthodes utilisées dans le domaine de la fiabilité des structures. Nous détaillons les méthodes d'approximation et les méthodes de simulation de Monte Carlo. Il est précisé quand ces dernières peuvent s'appliquer au calcul d'une espérance mathématique.

Le deuxième chapitre est entièrement consacré au tirage d'importance, qui est une méthode de Monte Carlo de réduction de variance, dont la principale difficulté est le choix de la densité de probabilité d'importance. Différentes approches existant en fiabilité des structures sont exposées. Nous nous intéressons ensuite au tirage d'importance adaptatif dans le cas paramétrique. L'idée est alors de construire une suite de densité qui convergerait vers la densité optimale. Nous y retrouvons aussi quelques illustrations.

Le troisième chapitre expose brièvement une méthode de simulation multiniveaux qui utilise les algorithmes de Monte Carlo par chaînes de Markov. Quelques exemples sont également présents.

Le quatrième chapitre est dédié à l'application industrielle. Le combustible à particules, les codes de calculs (ATLAS en particulier) simulant son comportement sous irradiation y sont présentés. Nous réalisons aussi une analyse de sensibilité du code ATLAS. Et enfin, nous appliquons les méthodes du tirage d'importance pour estimer la probabilité de rupture du combustible à particules.

Le cinquième chapitre de cette thèse a été fait avec Josselin Garnier et Bertrand Iooss dans le cadre du CEMRAC'S en 2006. Il concerne l'estimation de quantile (non extrême) avec peu de simulations. Les méthodes usuelles d'estimation de quantile sont présentées. Et nous proposons des méthodes qui exploitent l'information d'un modèle réduit pour estimer le quantile d'un modèle très coûteux. Ces méthodes sont testées sur une application industrielle.

## CHAPITRE 1

# MÉTHODES NUMÉRIQUES D'INTÉGRATION EN FIABILITÉ DES STRUCTURES

Nous commençons par présenter le formalisme utilisé dans le domaine de la fiabilité des structures. Soit  $X$  le vecteur aléatoire regroupant les paramètres d'entrée à valeurs dans  $\mathbb{R}^d$  de densité de probabilité  $f_X$  donnée. Pour chaque mode de défaillance de la structure, on définit une fonction d'état limite  $G$  qui est définie dans l'espace des paramètres du modèle. Le domaine de défaillance est défini par  $D_f = \{G(x) \leq 0, x \in \mathbb{R}^d\}$ , et le domaine de bon fonctionnement (ou domaine de sûreté) par  $D_s = \{G(x) > 0, x \in \mathbb{R}^d\}$ ; la frontière  $\{G(x) = 0\}$  est la surface d'état limite. La probabilité de défaillance  $P_f$  est donnée par :

$$P_f = \int \mathbf{1}_{\{G(x) \leq 0\}} f_X(x) dx = \mathbb{E} [\mathbf{1}_{\{G(X) \leq 0\}}]. \quad (1.0.1)$$

Pour l'évaluer, il faut donc calculer une intégrale, les méthodes d'intégration directe ne peuvent être menées que dans des cas particuliers favorables où  $G$  et les fonctions de densités ont des formes simples. Les méthodes de quadrature usuelles (méthodes trapézoïdales, de Simpson, de Gauss), sont souvent prohibitives à cause de la grande dimension de l'intégrale, et en plus, en fiabilité des structures la fonction à intégrer est discontinue. En outre la valeur de la probabilité de défaillance, en principe faible, conduit à des erreurs importantes car  $P_f$  peut-être de l'ordre de grandeur de l'erreur d'intégration. Le calcul de  $P_f$  est alors envisagé par des méthodes de simulations (simulations de Monte Carlo) ou un calcul analytique approché (méthodes d'approximation de type FORM-SORM).

Le présent chapitre contient un état de l'art des méthodes existant dans le domaine de la fiabilité des structures. Nous précisons aussi les conditions sous lesquelles ces méthodes sont applicables dans le contexte de calcul d'une espérance mathématique. Dans la première section, un point sur la modélisation de l'incertitude des paramètres d'entrée est présenté; dans la deuxième section, nous nous intéressons aux méthodes d'approximation et aux indices de fiabilité et dans la dernière section, nous rappelons le principe des simulations de Monte Carlo et nous passons en revue les méthodes dites de réduction de variance.

## 1.1 Traitement des incertitudes

Cette section a pour but de montrer comment peut s'opérer le traitement statistique en amont de toute analyse de fiabilité, ce qui permettra, d'une part de fixer le cadre des hypothèses statistiques et d'autre part d'identifier les différentes sources d'incertitudes intervenant dans le modèle probabiliste. Nous aborderons très brièvement les notions d'incertitude de modèle et d'analyse de sensibilité. Pour une revue détaillée de l'analyse de sensibilité, on pourra se référer au livre de A. Saltelli [SCS 00] ou à la thèse de J. Jacques [Jac 06], qui étudie notamment l'influence d'une incertitude de modèle dans une analyse de sensibilité.

### 1.1.1 Modèle et incertitudes

L'objet de la modélisation des incertitudes est de construire des modèles explicatifs et prévisionnels des phénomènes physiques et des modèles théoriques. Il existe plusieurs sources d'incertitudes :

1. les incertitudes épistémiques. Elles viennent du passage du phénomène réel au modèle mathématique. Elles sont dues à la modélisation mathématique et numérique du comportement de la structure (utilisation d'un code élément fini, de lois de comportement). Ces incertitudes peuvent être réduites par ajout d'informations (calculs ou expériences supplémentaires) ;
2. les incertitudes aléatoires. Elles apparaissent lors de l'estimation des paramètres du modèle. Elles sont dues à la représentation probabiliste des variables influençant le comportement de la structure, *i.e.* l'utilisation de variables aléatoires et de densité de probabilité caractérisant leurs lois ;

On suppose disposer d'échantillons des variables d'entrée issues par exemple de la fabrication, de collectes. On souhaite connaître les caractéristiques de la population infinie dont ils sont issues. Deux voies existent, l'approche inférentielle classique et l'approche bayésienne.

#### 1.1.1.1 Approche inférentielle classique

La statistique inférentielle classique suppose qu'une variable aléatoire appartient à une famille de lois paramétriques définies par leur densité par rapport à une mesure de référence  $\{f(x; \theta), \theta \in \Theta\}$ . Le vecteur des paramètres  $\theta$  est estimé d'après un échantillon d'observations (par la méthode du maximum vraisemblance, méthode des moments, ...). Une estimation précise nécessite l'emploi d'échantillons de grandes tailles. L'analyse théorique de la qualité de ces estimateurs peut reposer sur des théorèmes limites. La construction d'un intervalle de confiance

pour les paramètres  $\theta$  permet d'évaluer la précision des estimateurs. Du reste, l'approche classique, par l'intermédiaire de la théorie des tests, permet de prendre des décisions, de ne pas rejeter une hypothèse de modélisation statistique avec un niveau de confiance fixé (valider ou réfuter le choix d'une loi par exemple).

### 1.1.1.2 Approche bayésienne

La statistique Bayésienne suppose en plus qu'il est possible d'avoir une intuition sur les paramètres  $\theta$  sans pour autant faire appel aux données disponibles. Ces intuitions peuvent alors être traduites en terme de loi de probabilité sur  $\theta$ ; ce sont les lois a priori. Ces lois a priori sont définies par des paramètres, appelés hyperparamètres et devant être estimés sans avoir recours aux données disponibles. Par ce principe, la statistique Bayésienne se détache de la statistique classique puisque les paramètres du modèle, *i.e.*  $\theta$ , ne sont plus supposés constants mais traités comme des variables aléatoires. Toute la théorie Bayésienne repose sur un résultat simple mais qui permet de passer d'une loi a priori sur les paramètres (déterminée souvent à partir de jugements d'experts ou de retours d'expériences) à une loi dite a posteriori sur les paramètres sachant les données, c'est le théorème de Bayes :

$$f(\theta|x_1, \dots, x_n) = \frac{f(\theta)f_X(x_1, \dots, x_n|\theta)}{\int_{\Theta} f(\theta)f_X(x_1, \dots, x_n|\theta)d\theta}$$

où  $f(\theta|x_1, \dots, x_n)$  est la loi a posteriori,  $f(\theta)$  la loi a priori et  $f_X(x_1, \dots, x_n|\theta)$  la loi de  $X$  paramétrée par  $\theta$ . Dans le cas d'observation indépendantes :  $f_X(x_1, \dots, x_n|\theta) = \prod_{i=1}^n f_X(x_i, \theta)$ .

Ainsi, nous n'obtenons plus une estimation des paramètres mais une distribution pour ces paramètres.

Nous pouvons aussi prédire la densité de la variable aléatoire  $X$  au vue des réalisations d'un échantillon grâce à la formule suivante :

$$f_{X|x_1, \dots, x_n}(x) = \int_{\Theta} f_X(x_1, \dots, x_n|\theta)f(\theta|x_1, \dots, x_n)d\theta$$

### 1.1.2 Analyse de sensibilité

Procéder à une analyse de sensibilité d'un code de calcul est une étape essentielle, notamment dans le domaine de la fiabilité. Les objectifs les plus classiques de ces analyses sont les suivants :

- Prioritisation des variables : classer les variables par ordre d'importance ; si un facteur est influent par manque d'information, il peut être bon d'améliorer la connaissance de ce facteur (recherche d'informations, campagnes de mesures, ...) . En

fiabilité, la méconnaissance d'une variable d'entrée très influente peut avoir des répercussions énormes sur la probabilité de défaillance.

- Réduction du nombre de variables : les variables sont classées selon deux catégories, celles influentes et celles non influentes. Le modèle peut alors être simplifié aux premières, les secondes pouvant être considérées comme constantes. Il est alors possible d'obtenir une amélioration de la réponse en agissant sur le moins de variables possibles ; le fait de n'utiliser qu'un groupe de variables peut alléger le modèle fiabiliste choisi, et améliorer son efficacité ;
- Pouvoir discriminant : déterminer l'ensemble des variables responsables de la présence de la réponse dans une certaine région.

Un point plus précis sur les analyses de sensibilité sera fait dans le chapitre applicatif.

## 1.2 Indices de fiabilité et méthodes d'approximation

Cette section a pour objectif de présenter d'une part les indices de fiabilité et le lien qu'ils peuvent avoir avec la notion de probabilité de défaillance, et d'autre part les méthodes dites fiabilistes FORM et SORM, respectivement *First Order Reliability Method* et *Second Order Reliability Method*.

### 1.2.1 Indices de fiabilité

Les indices de fiabilité [MKL 86] sont des outils importants de la fiabilité des structures. Ils permettent de comparer entre elles les fiabilités de différentes structures. Un indice de fiabilité élevé indique une forte fiabilité et donc une faible probabilité de défaillance. L'historique et l'évolution de ces indices sont présentées dans [DM 96].

#### Indice de Rjanitzyne-Cornell

C'est le premier proposé dans la littérature [Cor 69], il est adapté à une surface de défaillance hyperplane, ce qui est un défaut notable. Il dépend de la manière dont la problème est formulé, c'est-à-dire de l'expression de la fonction d'état limite  $G$ . Il est défini comme le rapport entre l'espérance de la distribution de  $G(X)$  et de son écart-type :

$$\beta_C = \frac{\mathbb{E}[G(X)]}{\sqrt{\text{Var}(G(X))}}, \quad (1.2.1)$$

Parfois, il peut exister une relation directe entre la probabilité de défaillance et cet indice. Par exemple, prenons le cas particulier où l'on peut écrire  $G$  comme la différence entre une résistance et une sollicitation  $G(R, S) = R - S$  (voir figure (1.2.1)), et supposons que  $R$  et  $S$

soient distribuées suivant des lois normales indépendantes de moyennes respectives  $\mu_R$  et  $\mu_S$  et de variances respectives  $\sigma_R^2$  et  $\sigma_S^2$ . L'indice de Rjanitzyne-Cornell vaut :

$$\beta_C = \frac{\mu_R - \mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}},$$

et la probabilité de rupture s'exprime par :

$$P_f = P(G(R, S) \leq 0) = \Phi(-\beta_C),$$

où  $\Phi$  représente la fonction de répartition de la loi normale centrée réduite.

Il existe une extension de cet indice pour les fonctions d'état limite non linéaires basée sur un développement de Taylor du premier ordre et sur la connaissance de  $\mathbb{E}[G(X)]$  et de  $\text{Var}(G(X))$  [Mel 99], c'est ce qui est à la base de la méthode FORM.

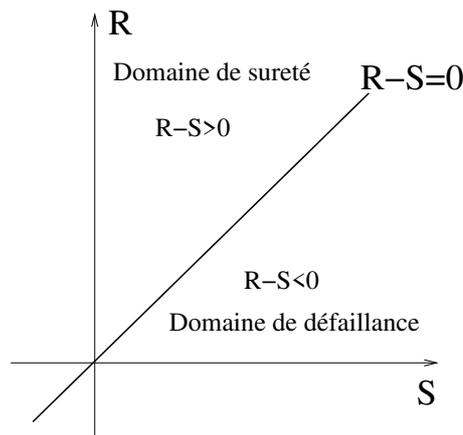


FIG. 1.2.1 – Illustration du problème  $R - S$

**Indice de Hasofer-Lind** C'est un indice géométrique exact et invariant par rapport à la formulation de  $G$  mais il a le défaut d'être indépendant par rapport à la géométrie de la surface d'état limite. Il est défini dans un espace gaussien. La transformation de l'espace des variables de base en un espace gaussien est brièvement présentée dans 1.2.2.1. Il correspond à la plus petite distance de l'origine du nouvel espace à la surface de défaillance :

$$\beta_{HL} = \min_{H(u)=0} (u^t u)^{1/2} \quad (1.2.2)$$

où  $H$  est la surface d'état limite dans le nouvel espace.

Le ou les points réalisant le minimum de (1.2.2) sont appelés points de conception.

Dans les cas où la fonction d'état limite est un hyperplan, l'indice de Hasofer-Lind est égal à l'indice de Rjanitzyne-Cornell. Reprenons le cas particulier précédent  $G(R, S) = R - S$ . En

transformant les variables  $R$  et  $S$  dans l'espace gaussien, on obtient les nouvelles variables normales centrées réduites  $R_u$  et  $S_u$  :

$$R_u = \frac{R - \mu_R}{\sigma_R}, \quad S_u = \frac{S - \mu_S}{\sigma_S}.$$

La surface d'état limite  $H$  dans le nouvel espace est alors :

$$H(R_u, S_u) = \mu_R - \mu_S + \sigma_R R_u - \sigma_S S_u.$$

Les coordonnées du point de la surface d'état limite  $H(R_u, S_u) = 0$  le plus proche de l'origine de l'espace gaussien sont :

$$\left( \frac{S_u - R_u}{\sigma_R^2 + \sigma_S^2} \sigma_R, \frac{R_u - S_u}{\sigma_R^2 + \sigma_S^2} \sigma_S \right), \quad (1.2.3)$$

et l'indice de Hasofer-Lind vaut :

$$\beta_{HL} = \frac{\mu_R - \mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}}.$$

### Indice de fiabilité généralisé

Cet indice a été introduit par Ditlezen [Dit 79], il dépend de la probabilité de défaillance :

$$\beta_G = \Phi^{-1}(1 - P_f). \quad (1.2.4)$$

Il est estimé par :

$$\hat{\beta}_G = \Phi^{-1}(1 - \hat{P}_f), \quad (1.2.5)$$

où  $\hat{P}_f$  est une estimation de la probabilité de défaillance (calculée par FORM par exemple). Cet indice est égal aux indices de Hasofer-Lind et Rjanitzyne-Cornell si la surface de défaillance est un hyperplan dans l'espace gaussien. Cet indice requiert malheureusement la connaissance de la probabilité de défaillance, qui est ce que l'on cherche.

Dans le cas particulier  $G(R, S) = R - S$ ,  $R$  et  $S$  distribuées suivant des lois normales indépendantes de moyennes respectives  $\mu_R$  et  $\mu_S$  et de variances respectives  $\sigma_R^2$  et  $\sigma_S^2$ , la probabilité de défaillance est donnée par :

$$P_f = P(R - S \leq 0) = \Phi \left( -\frac{\mu_R - \mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}} \right),$$

et donc l'indice généralisé vaut :

$$\beta_G = \frac{\mu_R - \mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}}.$$

### 1.2.2 Méthodes FORM/SORM

Ces méthodes procèdent en quatre étapes distinctes que nous détaillerons un peu dans cette partie :

1. la première consiste en un changement de variables des données d'entrées pour se ramener à des variables aléatoires normales centrées réduites et indépendantes ;
2. la seconde préconise la recherche du ou des points de conception (points de défaillance les plus probables), c'est-à-dire le ou les points sur la surface d'état limite les plus proches de l'origine :
3. la troisième étape consiste à approximer la surface de défaillance, soit par un hyperplan (FORM) soit par une surface quadratique (SORM) ;
4. la dernière étape nous permet de calculer la probabilité de défaillance à partir de la surface d'approximation.

#### 1.2.2.1 Changement de l'espace d'entrée des variables

On utilise une transformation  $T : x \rightarrow u$ , qui a un vecteur aléatoire  $X$  de dimension  $d$  et de densité  $f$  quelconque associe un vecteur aléatoire  $U$  de densité  $\varphi_d$  multinormale centrée réduite de dimension  $d$ . La figure 1.2.2 illustre ce changement d'espace d'entrée des variables en dimension 2. La fonction de défaillance et la probabilité de défaillance s'écrivent alors :

$$H(u) = G(T^{-1}(u)), \text{ et } P_f = \int_{H(u) \leq 0} \varphi_d(u) du.$$

La transformation de Rosenblatt [Ros 52], [HR 81] est utilisée lorsque l'on connaît la densité conjointe des variables. Dans le cas où les variables sont corrélées, cette transformation n'est pas unique car elle dépend de l'ordre dans lequel les variables sont traitées. Si on connaît seulement les densités marginales, les deux premiers moments et les corrélations, l'utilisation de la transformation de Nataf est préconisée [Nat 62][LDK 86]. Si l'influence des moments d'ordre supérieur à deux est recherchée, la transformation d'Hermite (décomposition sur une base de fonctions gaussiennes) peut-être utilisée sans exiger le choix d'une densité de probabilité. L'influence du choix de la transformation est discutée dans [Dev 96] et [DM 96], il apparaît que celle-ci est limitée.

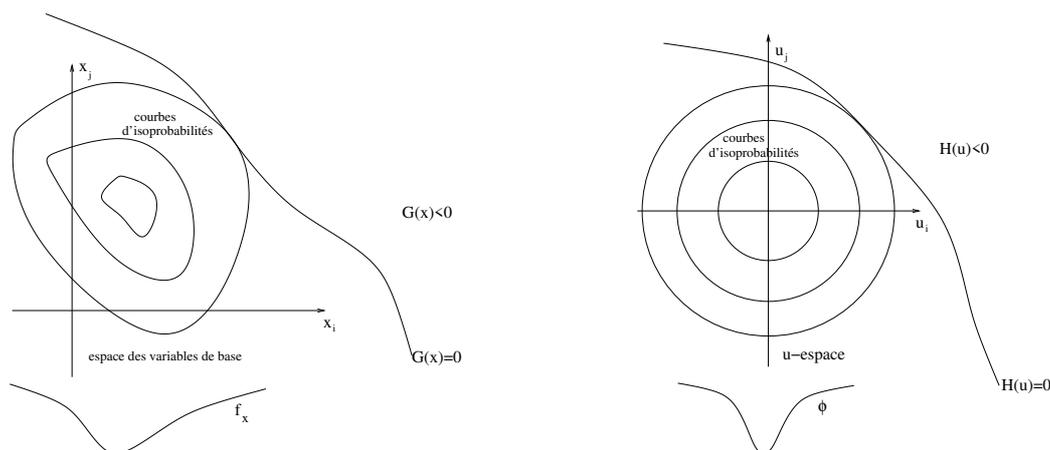


FIG. 1.2.2 – Transformation de l'espace d'entrée des variables

### 1.2.2.2 Recherche des points de conception

La recherche du point de conception ou du point de défaillance le plus probable relève d'un problème d'optimisation non linéaire sous contrainte dans l'espace gaussien.

$$(\mathcal{P}) \begin{cases} \text{trouver } u^* \text{ qui minimise } u^t u \\ \text{sous la contrainte } H(u) = 0 \end{cases} \quad (1.2.6)$$

C'est une étape cruciale du calcul de la probabilité de défaillance par FORM/SORM. Il faut utiliser un algorithme qui converge le plus rapidement possible (minimiser le nombre de simulations faisant intervenir le calcul de la fonction d'état limite) et qui rend compte de tous les points de conception. Dans ce problème d'optimisation, la fonction à minimiser est d'une forme simple, quadratique et convexe. Ainsi, c'est la forme des contraintes qui déterminera l'algorithme le plus approprié.

### Algorithmes de recherche locale

On peut classer les algorithmes de recherche locale en trois catégories :

1. Les algorithmes ne faisant pas intervenir le calcul du gradient de  $H$  comme les méthodes de dichotomie ou du simplexe. Elles sont intéressantes lorsque le calcul du gradient est délicat, mais impose une vitesse de convergence lente. Ces algorithmes ne répondent pas au problème de recherche des points de conception.
2. Les algorithmes d'ordre un qui nécessite le calcul du gradient de  $H$ , comme la méthode du gradient projeté ou l'algorithme HL-RF, présenté ultérieurement ; ce dernier a été conçu pour répondre spécifiquement au problème de recherche des points de conception.

3. Les algorithmes d'ordre deux, comme les méthodes de Newton ou les méthodes de programmation quadratique séquentielle (SQP).

Si les fonctions objective et contrainte sont suffisamment régulières, les algorithmes basés sur le gradient sont les plus efficaces. Il est souhaitable, autant que possible, d'utiliser une expression analytique des gradients; cependant, celle-ci n'est pas toujours disponible et le gradient est alors approché par la méthode des différences finies, qui peut être coûteuse en temps de calcul et peut même conduire à une perte de convergence si les gradients sont mal estimés.

Il est possible d'utiliser les conditions d'optimalité de Karush-Kuhn-Tucker. Soit le lagrangien du problème  $L(u, \lambda) = \|u\| + \lambda H(u)$ ,  $u^*$  est un extremum local de  $(\mathcal{P})$  si et seulement si il existe un réel positif  $\lambda^*$  tel que :

$$\nabla L(u^*, \lambda^*) = u^{*t} + \lambda^* \nabla H(u^*) = 0 \text{ et } H(u^*) = 0.$$

La recherche de  $u^*$  se fait itérativement :  $u^{(k+1)} = u^{(k)} + \alpha^{(k)} d^{(k)}$  avec  $\alpha^{(k)} \in \mathbb{R}^+$  et  $|d^{(k)}| = 1$ . Les différences entre les algorithmes se situent au niveau de la direction de descente  $d^{(k)}$  et du pas de descente  $\alpha^{(k)}$ .

#### *Algorithme HL-RF*

Cet algorithme a été initié par Hasofer et Lind [HL 74] puis étendu par Rackwitz et Fiessler [RF 78]. Il est spécifique à la résolution du problème  $(\mathcal{P})$ . Il ne nécessite qu'un nombre limité de calculs à chaque itération.

$$u^{(k+1)} = \left( u^{(k)t} \beta^{(k)} \right) \beta^{(k)} + \frac{H(u^{(k)})}{\|\nabla H(u^{(k)})\|} \beta^{(k)} \text{ avec } \beta^{(k)} = -\frac{\nabla H(u^{(k)})}{\|\nabla H(u^{(k)})\|}. \quad (1.2.7)$$

La convergence globale de cet algorithme n'est pas assurée. Pour palier cette difficulté, Zhang et Der Kiureghian ont proposé une amélioration de HL-RF dans [ZDK 94] en ajustant le pas de descente à chaque itération.

#### *Algorithmes de type SQP*

Les algorithmes de programmation quadratique séquentielle utilisent une approximation quadratique du lagrangien. Ils nécessitent le calcul du hessien de la fonction d'état limite à chaque itération ce qui est lourd et coûteux en grande dimension, c'est pourquoi ces méthodes sont surtout efficaces (robustesse et vitesse de convergence) en petite dimension. L'algorithme Abdo-Rackwitz est une alternative à la méthode SQP pour les problèmes en grande dimension. Cet algorithme adapte la méthode SQP à la forme particulière de la fonction objective; comme l'algorithme HL-RF, il ne s'applique qu'à la résolution du problème  $(\mathcal{P})$ .

## Algorithmes de recherche globale

Pour résoudre un problème d'optimisation globale, il est possible de recourir aux algorithmes d'optimisation stochastique comme le recuit simulé ou les algorithmes génétiques [Duf 96]. Dans [DKD 98], Der Kiureghian et Dakessian proposent une méthode pour forcer la convergence de l'algorithme d'optimisation vers un autre point de conception que celui déjà trouvé. Si  $u_i^*$  est un point de conception déjà connu, on exclut le voisinage de ce point de conception de l'ensemble des solutions admissibles par l'ajout d'une fonction barrière  $B_i$  positive. On perturbe ainsi tous les points de conception déjà trouvés afin de déterminer le suivant.

Soient  $u_i^*, i = 1, \dots, p$  les  $p$  premiers points de conception trouvés. L'expression de la surface d'état limite est remplacée par :  $H_p(u) = H(u) + \sum_{i=1}^p B_i(u)$ , avec

$$B_i(u) = \begin{cases} s_i (r_i^2 - \|u - u_i^*\|^2)^2 & \text{si } \|u - u_i^*\| \leq r_i \\ 0 & \text{sinon} \end{cases} \quad (1.2.8)$$

où  $s_i$  et  $r_i$  sont des paramètres positifs définissant les « barrières »  $B_i$ .

L'algorithme HL-RF est alors appliqué au problème d'optimisation sous une nouvelle contrainte (nouvelle expression de la fonction d'état limite). Cet algorithme permet de ne pas converger à nouveau vers un point de conception déjà trouvé, cependant, il peut converger vers un optimum artificiel. Dans [DKD 98], il est précisé que cette situation indique qu'il n'existe alors pas d'autres points de conception que ceux déjà identifiés. On peut s'en servir en tant que critère d'arrêt.

### 1.2.2.3 Approximation de la surface d'état limite

Dans les méthodes FORM et SORM, la grande différence provient du choix de l'approximation de la surface d'état limite. Une illustration de ces méthodes est donnée figure (1.2.3).

#### Approximation FORM

On remplace la surface d'état limite par son hyperplan tangent au point de conception, c'est-à-dire que l'on effectue un développement de Taylor d'ordre un au point de conception. La linéarisation de la surface d'état limite peut induire une perte de précision due à la géométrie du domaine. Pour une surface d'état limite convexe au voisinage du point de conception, la méthode FORM sous-estime la probabilité de défaillance, et dans le cas d'une surface d'état limite concave, elle la surestime. Cependant, cette méthode est peu coûteuse et peut être efficace si la surface d'état limite est assez lisse.

L'approximation linéaire de la surface d'état limite au point de conception s'écrit :

$$\nabla H(u)^t |_{u=u^*} (u - u^*) = 0, \quad (1.2.9)$$

Elle peut aussi s'écrire :

$$a^t u + \beta_{HL} = 0, \quad a = \frac{\nabla H(u) |_{u=u^*}}{\|\nabla H(u) |_{u=u^*}\|}, \quad (1.2.10)$$

et l'indice de fiabilité de Hasofer-Lind a une expression simplifiée :

$$\beta_{HL} = -\frac{\nabla H(u) |_{u=u^*}}{\|\nabla H(u) |_{u=u^*}\|} \cdot u^*.$$

S'il existe plusieurs points de conception, on traite alors un problème de défaillance en série (approche multi-FORM), l'union des domaines de défaillances approchés par linéarisation de la surface de défaillance en chacun des minima devient alors le domaine de défaillance sur lequel sera calculée la probabilité de rupture.

### Approximation SORM

Ici, la surface d'état limite est approchée par une surface quadratique, soit par l'intermédiaire d'un développement de Taylor au second ordre au point de conception [Tve 90] (ajustement de courbures), soit par l'utilisation d'une surface parabolique avec des points appartenant à la surface d'état limite [LDKH 87], [ZDK 94] (ajustement par points). L'approximation par ajustement de courbures nécessite le calcul du Hessien, qui peut-être coûteux en temps de calcul.

L'approximation quadratique de la surface d'état limite au point de conception, par ajustement de courbures s'écrit :

$$\nabla H(u)^t |_{u=u^*} (u - u^*) + \frac{1}{2} (u - u^*)^t \frac{\partial^2 H(u)}{\partial u_i \partial u_j} |_{u=u^*} (u - u^*) = 0, \quad (1.2.11)$$

La surface d'état limite peut aussi être approchée par régression, on construit une surface de réponse de la fonction d'état limite, voir la section 1.2.3.

Si il existe plusieurs points de conception, la surface de défaillance est approchée par une surface quadratique en chacun des minima, la formule de Breitung (1.2.13) nous donne une approximation de la probabilité pour chaque minima ; un hyperplan équivalent est alors obtenu à partir de l'indice généralisé (1.2.4) de chaque approximation. On peut alors traiter le problème comme un problème multi-FORM.

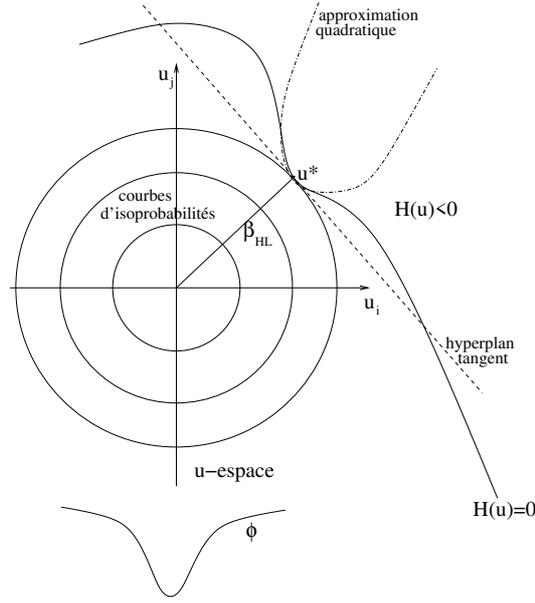


FIG. 1.2.3 – Méthodes FORM et SORM

#### 1.2.2.4 Calcul de la probabilité de défaillance

Pour la méthode FORM, on se sert de l'indice de fiabilité de Hasofer-Lind donné par (1.2.2), la probabilité de défaillance est approchée par :

$$P_f \approx 1 - \Phi(\beta_{HL}) \quad (1.2.12)$$

où  $\Phi$  est la fonction de répartition d'une loi normale centrée réduite.

Dans le cas d'un calcul multi-FORM avec  $k$  points de conception, on exclue les zones de recouvrement et la probabilité de défaillance est approchée par :

$$P_f \approx \sum_{i=1}^k \left( 1 - \Phi(\beta_{HL}^{(i)}) \right) - \sum_{i=1}^k \sum_{j=1}^{i-1} F_2 \left( -\beta_{HL}^{(i)}, -\beta_{HL}^{(j)}, \rho_{ij} \right),$$

où  $F_2 \left( -\beta_{HL}^{(i)}, -\beta_{HL}^{(j)}, \rho_{ij} \right)$  est la fonction de répartition d'un vecteur de deux lois normales centrées réduites et de coefficient de corrélation  $\rho_{ij}$  au point  $(-\beta_{HL}^{(i)}, -\beta_{HL}^{(j)})$ .

Souvent, la méthode SORM est une aide à la validation d'un calcul FORM, c'est-à-dire qu'elle peut servir à vérifier le caractère linéaire d'une surface d'état limite. La formule la plus utilisée pour la méthode SORM est celle de Breitung [Bre 84] :

$$P_f \approx \Phi(-\beta_{HL}) \prod_{i=1}^{n-1} \frac{1}{\sqrt{1 + \beta_{HL} k_i}} \quad (1.2.13)$$

où  $k_i$  représente les courbures principales de la surface d'état limite au point de conception  $u^*$ . Elle est valable lorsque  $\beta_{HL} \rightarrow +\infty$ . Le calcul des courbures principales nécessite le calcul

de la matrice hessienne de la fonction d'état limite au point de conception et le calcul des valeurs propres d'une matrice de taille  $(d-1) \times (d-1)$ . Si les dérivées partielles de la fonction d'état limite n'ont pas de forme explicite, l'approximation de la matrice hessienne (différences finies par exemple) est irréalisable en grande dimension car beaucoup trop coûteuse. Il est alors nécessaire d'approcher la surface d'état limite par une fonction ayant une forme explicite et deux fois différentiable au point de conception.

D'autres formules approchées de la probabilité de défaillance ont été développées depuis, on peut les retrouver dans : [HR 88] (celle-ci ressemble fortement à celle de Breitung), [CE 94], [ZO 99a], [Hon 99] ou [PBP 99]. Pour certaines configurations, l'utilisation de ces formules peuvent conduire à de mauvaises approximations. En outre, certains auteurs se sont intéressés au calcul exact d'intégration d'une surface quadratique dans l'espace gaussien [Hel 83, Ric 80, Tve 90]. Dans le même style, on retrouve la méthode RGMR (Riemannian Geometrical Methods for Reliability) développée dans [Mit 99] et [MBL 95]. Une autre méthode, basée sur l'inverse de la transformée de Fourier rapide est présentée dans [ZO 99b].

### 1.2.2.5 Fiabilité des systèmes

Cette section s'inspire de [CDP 05]. On parle de systèmes, soit quand on dispose d'un ensemble de composants dépendant les uns des autres, soit quand on dispose d'un seul composant présentant plusieurs modes de défaillances.

La démarche pratique courante est la suivante :

1. identification des composants et/ou des modes de défaillances, et de leurs interactions éventuelles ;
2. représentation du problème en un problème type combinaison série ou parallèle (arbres de défaillance) ;
3. évaluation de la probabilité de chaque mode de défaillance de chaque composant ;
4. évaluation de la probabilité du système.

### Combinaison de modes de défaillance

#### *Système série*

Un système, composé de  $m$  modes de défaillances, est un système série si l'occurrence d'un seul événement entraîne la défaillance totale du système. A chaque mode de défaillance  $m$  correspond une surface d'état limite  $\{G_m(x) \leq 0\}$ .  $P_f$  est alors la probabilité de l'union des

modes de défaillance :

$$P_f = P \left( \bigcup_{i=1}^m G_i(X) \leq 0 \right)$$

Pour l'approximation FORM, , on peut définir une surface d'état limite (formule (1.2.10)) pour chaque mode de défaillance  $m$

$$a_m^t u + \beta^i = 0, \quad a_m = \frac{\nabla H_m(u)|_{u=u_m^*}}{\|\nabla H_m(u)|_{u=u_m^*}\|},$$

où  $u_m^*$ , est le point de conception pour le mode  $m$  de défaillance et  $H_m$  la fonction d'état limite associée à ce mode. L'estimateur de la probabilité de défaillance est alors donné par :

$$\hat{P}_f = 1 - \Phi_m(\beta, R), \quad (1.2.14)$$

où  $\Phi_m(\beta, R)$  est la fonction de répartition d'une loi multinormale centrée de dimension  $m$  de matrice de corrélation  $R$  ( $r_{ij} = a_i^t a_j$ ,  $r_{ii} = 1$ ) et évaluée en  $\beta = (\beta_{HL}^1, \dots, \beta_{HL}^m)^t$ .

#### *Système parallèle*

Un système, composé de  $m$  modes de défaillances est un système parallèle si la défaillance de tous de les événements est nécessaire à la défaillance du système. A chaque mode de défaillance  $m$  correspond une surface d'état limite  $\{G_m(x) \leq 0\}$ .  $P_f$  est alors la probabilité de l'intersection des modes de défaillance :

$$P_f = P \left( \bigcap_{i=1}^m G_i(X) \leq 0 \right)$$

De la même manière que pour un système série, on peut approcher la probabilité de défaillance, pour la méthode FORM, par :

$$\hat{P}_f = \Phi_m(-\beta, R). \quad (1.2.15)$$

où  $\Phi_m(\beta, R)$  est la fonction de répartition d'une loi multinormale centrée de dimension  $m$  de matrice de corrélation  $R$  ( $r_{ij} = a_i^t a_j$ ) et évaluée en  $\beta = (\beta_{HL}^1, \dots, \beta_{HL}^m)^t$ .

Ces approximations de la probabilité de défaillance par des fonctions de répartition de loi multinormale sont dues à [HR 83].

Pour des informations détaillées sur la fiabilité des systèmes, on peut se reporter aux livres de M. Lemaire [Lem 05] et de R. E. Melchers [Mel 99].

#### **1.2.2.6 Quelques remarques**

Ces méthodes d'approximation sont efficaces car le nombre de simulations requis ne dépend pas du niveau de probabilité cherché (formule de Breitung (1.2.13)), de plus elles permettent d'effectuer des analyses de sensibilité [DM 96].

En revanche, elles ne permettent pas une estimation de l'erreur et impose des conditions de régularité et différentiabilité de la fonction d'état limite  $H$ . En définitive, les véritables limites de ces méthodes interviennent lorsque la fonction d'état limite est de type oscillatoire et/ou lorsqu'il existe plusieurs points de conception. Même lorsque les différents points de conception sont trouvés, un problème de combinaison des différentes approximations survient, l'approche multi-FORM ([Dev 96], [DM 96]), basée sur la formule de Poincaré peut être utilisée. Mais si les différents points de conception sont très proches et/ou si le domaine de défaillance est constitué d'un ensemble de sous domaines disjoints, il est très difficile d'obtenir des approximations correctes de la probabilité de défaillance [DKD 98].

### 1.2.3 Couplage mécano-fiabiliste

Pour une présentation détaillée des problèmes de couplages, on peut se reporter au livre de M. Lemaire [Lem 05]. Deux approches sont envisagées :

#### Le couplage direct

Le calcul de fiabilité fait directement appel au code mécanique (souvent un modèle utilisant les éléments finis) à chaque calcul de la fonction d'état limite. Pour une méthode fiabiliste choisie, chaque fois que l'algorithme requiert une évaluation mécanique, le logiciel fiabiliste envoie un jeu de paramètres au logiciel mécanique qui renvoie à son tour la réponse mécanique associée, qui servira à l'évaluation de la fonction d'état limite. La figure (1.2.4) illustre le couplage direct.

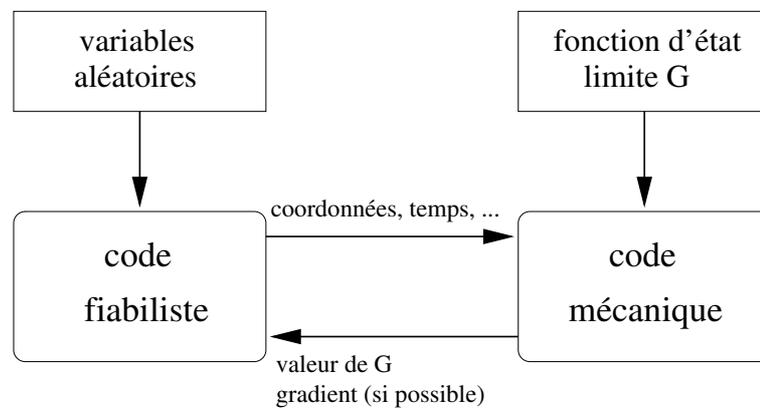


FIG. 1.2.4 – Principe du couplage direct (illustration issue de [CDP 05])

#### Le couplage par surface de réponse

On utilise un modèle réduit explicite soit pour approcher le code mécanique, soit pour approcher la surface d'état limite.

#### *Approximation du modèle mécanique*

On parle alors de stratégie de surface de réponse globale. La principale question concerne l'erreur associée à l'utilisation de cette surface de réponse, qui est une approximation du modèle mécanique, lui-même approximation de la réalité ; de plus, le nombre de simulations nécessaires à la construction d'un modèle valide globalement peut être élevé.

Toutes les familles de surfaces de réponse peuvent être utilisées dans cette approche, on peut citer les splines, les modèles linéaires généralisés, les modèles PLS (Partial Least Square), les machines à vecteurs support, ... Pour obtenir des informations complètes, on peut se référer à [HTF 02].

#### *Approximation de la surface d'état limite*

On parle ici de stratégie locale. Le problème d'optimisation ( $\mathcal{P}$ ) doit alors être résolu sous la contrainte  $\tilde{H}(u) = 0$  où  $\tilde{H}$  est la surface de réponse approchant  $H$ . Pour une description détaillée, on peut consulter [Dev 96] ou [ML 99]. Les avantages de l'utilisation des surfaces de réponse résultent principalement de la simplification des calculs et en particulier dans le cas où le calcul du gradient par différences finies n'est pas possible ou trop coûteux. Les principales variantes viennent des différents choix possibles de familles de fonctions d'approximations et des différents manières de construire cette fonction. Elles doivent aussi répondre à des besoins essentiels en fiabilité :

1. avoir une expression analytique qui évite de déformer les réponses dans les queues de distribution ;
2. avoir des formes facilement différentiables pour faciliter la recherche des points de conception.

Pour répondre à ces attentes, les fonctions polynomiales sont fréquemment utilisées, elles possèdent en effet des propriétés intéressantes (simplicité de calculs, différentiabilité, ...). La construction de ces fonctions peut être réalisée par interpolation, régression ou encore ajustements successifs de polynômes à une seule variable.

Des applications des méthodes de surface de réponse dans les problèmes de fiabilité ont déjà été testées par de nombreux auteurs [Far 89], [BB 90], [MLET 92], [MLBL 93], [RE 93], [EFS 94], [KN 97].

Bucher [BB 90] propose une méthode itérative de construction de surfaces de réponse dans le modèle physique à partir d'un plan d'expériences tirant profit du modèle probabiliste. Le centre du premier plan d'expérience  $\bar{x}^{(0)}$  est le point moyen, et les autres points sont choisis autour du point moyen et dans la direction des variables  $x_i$  à une distance  $k_i\sigma_{x_i}$  où  $k_i$  est une valeur arbitraire et  $\sigma_{x_i}$  l'écart-type de la variable  $x_i$ . A chaque itération,  $f_i$  est réduit et le nouveau point central  $\bar{x}^{(k+1)}$  est obtenu par :

$$\bar{x}^{(k+1)} = \bar{x}^{(k)} + \left( x^{(k)*} - \bar{x}^{(k)} \right) \frac{G(\bar{x}^{(k)})}{G(\bar{x}^{(k)}) - G(x^{(k)*})} \quad (1.2.16)$$

où  $x^{(k)*}$  est le point de conception calculé à l'itération  $k$ . Cette approche a été reprise et améliorée dans [RE 93] et [KN 97].

Des auteurs [MLET 92], [MLBL 93] et [EFS 94] proposent respectivement les méthodes SQR (Surface de Réponse Quadratique) et ARERSA ; ces méthodes sont basées sur le principe que l'on peut toujours construire une boule au voisinage du point de conception dans laquelle un polynôme de degré au plus deux est une bonne approximation.

Dans sa thèse, N. Devictor [Dev 96] propose un algorithme adaptatif de construction de surface de réponse nommé RSAED (Response Surface with Adaptive Experimental Design). Cet algorithme consiste à adapter les plans d'expérience successifs pour les centrer dans les régions les plus intéressantes à partir des résultats obtenus aux étapes précédentes. De plus il permet de réutiliser des points précédemment calculés, il permet aussi de perturber des points trop proches afin que chaque calcul apporte de l'information supplémentaire, et enfin d'utiliser de nombreux indicateurs pour tester la qualité du plan d'expérience et de la surface de réponse. On peut trouver une autre procédure adaptative dans [HH 99].

Dans sa thèse, M. Pendola [Pen 00] propose deux méthodes permettant d'obtenir des résultats d'un modèle mécanique complexe en utilisant un modèle simplifié (surface de réponse) et étudie l'impact d'un modèle simplifié sur la fiabilité. La première nommée Facteur de Correction de Modèle suppose que dans un voisinage du point de conception, le rapport des solutions des deux fonctions d'état limite peut être localement remplacé par une constante (en considérant que les deux modèles se comportent de la même manière). Cette méthode n'est pas applicable lorsque le modèle complexe fait appel aux éléments finis. La deuxième méthode, nommée Approximation Fonctionnelle de l'Écart repose sur la quantification de l'écart de la réponse fournie par les deux modèles, c'est-à-dire que l'écart de la réponse entre les deux modèles est supposé être une fonction déterministe des paramètres des modèles. En fait, plutôt que d'approcher toute la surface d'état limite, on approche seulement l'écart.

Des méthodes de surface de réponse par réseaux de neurones ont aussi été utilisés dans [PLT 98] et [DM 00], un réseau de neurones est substitué à la réponse du modèle.

Une méthode de surface de réponse stochastique est proposée dans [IRG 98] et [Ber 05], les polynômes de chaos sont utilisés comme approximation de la fonction d'état limite dans les analyses de fiabilité.

Récemment, les méthodes de krigeage ont aussi été proposées en tant que surface de réponse de la surface d'état limite ; cette approche reprend le principe de Bucher décrit par la formule (1.2.16) [Kay 05].

### 1.3 Méthodes de simulation

Les méthodes de simulations de Monte Carlo sont des méthodes générales d'estimation d'intégrales multidimensionnelles qui peuvent être écrites comme des espérances mathématiques ; elles peuvent aussi servir à la résolution d'équations aux dérivées partielles, à la résolution de systèmes linéaires et à la résolution de problèmes d'optimisation. Dans un contexte général, on cherche à estimer :

$$I = \int_{D \subset \mathbb{R}^d} g(x)f(x)dx, \quad (1.3.1)$$

où  $f$  est une densité de probabilité.

En fiabilité des structures  $g$  est une fonction indicatrice, qui est une fonction discontinue et l'intégrale cherchée est présentée par la formule (1.0.1).

#### 1.3.1 Monte Carlo direct

C'est la méthode la plus simple à mettre en oeuvre, elle consiste à tirer des réalisations du vecteur aléatoire  $X$  suivant leur loi de probabilité initiale.

##### 1.3.1.1 Contexte général

La méthode Monte Carlo directe consiste à écrire l'intégrale  $I$  sous la forme  $I = \mathbb{E}[g(X)]$  où  $X \sim f$  (*i.e.*  $X$  est une variable aléatoire de loi à densité  $f$ ) et à générer  $n$  variables aléatoires indépendantes et identiquement distribuées (*v.a.i.i.d.*) de densité  $f$  pour calculer  $I$ . En effet, si  $(X_i)_{1 \leq i \leq n}$  est une suite de *v.a.i.i.d.* de densité  $f$  et en supposant  $g(X) \in L_f^1(\mathbb{R}^d)$  (*i.e.*  $\mathbb{E}_f[|g(X)|] < +\infty$ ) par la *loi forte des grands nombres*, nous avons :

$$\hat{I}_n^{MC} = \frac{1}{n} \sum_{i=1}^n g(X_i) \xrightarrow[n \rightarrow +\infty]{} \mathbb{E}[g(X)] \quad p.s. \quad (1.3.2)$$

$\hat{I}_n^{MC}$  est un estimateur sans biais de  $I$ . Sa variance est donnée par :

$$\text{Var} \left( \hat{I}_n^{MC} \right) = \frac{1}{n} \text{Var} (g(X)).$$

Si  $\mathbb{E}_f[g(X)^2] < +\infty$  (i.e.  $\sigma^2 = \text{Var}_f(g(X)) < +\infty$ ) le *théorème central limite* donne la convergence suivante :

$$\sqrt{n}(I - \hat{I}_n^{MC}) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2), \quad (1.3.3)$$

et nous pouvons alors obtenir un intervalle de confiance de niveau  $1 - 2\alpha$  pour  $I$  :

$$\left[ \hat{I}_n^{MC} - \Phi^{-1}(\alpha) \sqrt{\frac{\widehat{\text{Var}}(g(X))}{n}}, \hat{I}_n^{MC} + \Phi^{-1}(\alpha) \sqrt{\frac{\widehat{\text{Var}}(g(X))}{n}} \right], \quad (1.3.4)$$

$$\text{où } \widehat{\text{Var}}(g(X)) = \frac{1}{n-1} \sum_{i=1}^n \left( g(X_i) - \hat{I}_n^{MC} \right)^2.$$

La vitesse de convergence de cette méthode est de l'ordre de  $1/\sqrt{n}$ . Cette vitesse présente l'avantage d'être insensible à la dimension et de ne pas dépendre de la régularité de la fonction  $g$  à intégrer ( $g(X_1)$  doit seulement être de carré intégrable). Une comparaison des vitesses de convergence des méthodes de quadrature classiques et de la méthode de Monte Carlo directe est donnée dans le tableau (1.3.1).

Méthode	Vitesse
Monte Carlo	$n^{-1/2}$
Règle trapézoïdale	$n^{-2/d}$
Règle de Simpson	$n^{-4/d}$
Règle de Gauss ( $m$ points)	$n^{-(2m-1)/d}$

TAB. 1.3.1 – *Vitesses de convergence en dimension  $d$  pour des fonctions régulières*

### 1.3.1.2 Fiabilité des structures

Dans ce paragraphe, la méthode de Monte Carlo directe pour l'estimation de probabilité de défaillance est brièvement vue.

La probabilité de rupture pour un domaine de défaillance donné est la moyenne empirique de la fonction indicatrice  $\mathbf{1}_D(\cdot)$  sur un échantillon  $\{X_1, \dots, X_n\}$  simulé indépendamment et identiquement suivant  $f$ . Nous cherchons à calculer la quantité suivante :

$$P_f = \int \mathbf{1}_D(x) f(x) dx$$

L'estimateur de cette intégrale est :

$$\hat{P}_f = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_D(X_i),$$

et sa variance est :

$$\text{Var}(\hat{P}_f) = \frac{P_f(1 - P_f)}{n}.$$

Un intervalle de confiance de niveau  $1 - 2\alpha$  pour  $P_f$  est donné par :

$$\left[ \hat{P}_f - \Phi^{-1}(\alpha) \sqrt{\frac{\hat{P}_f(1 - \hat{P}_f)}{n}}, \hat{P}_f + \Phi^{-1}(\alpha) \sqrt{\frac{\hat{P}_f(1 - \hat{P}_f)}{n}} \right]. \quad (1.3.5)$$

La figure (1.3.1) illustre le principe de la simulation directe dans un contexte de fiabilité des structures.

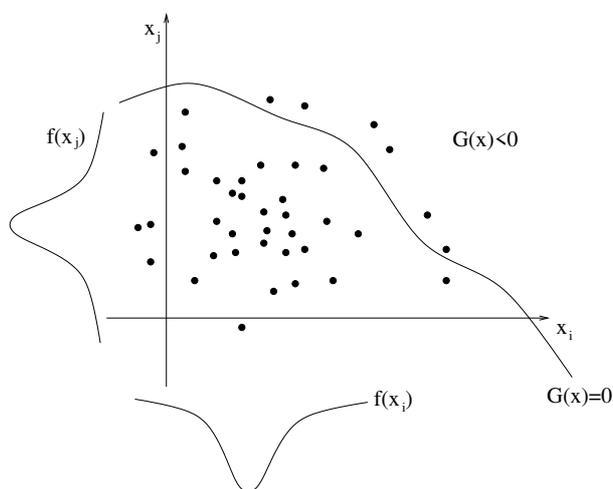


FIG. 1.3.1 – Monte Carlo direct

L'efficacité de cette méthode peut-être mesurée par le coefficient de variation  $\delta$  de l'estimateur, qui est défini comme le rapport entre l'écart-type et la moyenne :

$$\delta^2 = \frac{1 - P_f}{nP_f}.$$

L'expression du coefficient de variation dépend seulement de la probabilité de rupture et du nombre de simulations  $n$ .  $\delta$  peut-être estimé en remplaçant  $P_f$  par  $\hat{P}_f$ .

Quand  $P_f$  est petit, le nombre  $n_\delta$  de simulations nécessaires à l'obtention d'un coefficient de variation donné est inversement proportionnel à la probabilité de défaillance :

$$n_\delta = \frac{1 - P_f}{P_f \delta^2} \sim \frac{1}{P_f \delta^2}$$

Ainsi, pour obtenir une probabilité de rupture de  $10^{-3}$  avec un coefficient de variation de 20%, il faut 25000 simulations.

Si on cherche à estimer une probabilité de l'ordre de  $10^{-l}$ , il faut au moins effectuer  $10^{l+2}$  simulations pour obtenir un coefficient de variation de 10%.

Comme nous l'avons vu, l'obtention de résultats suffisamment précis à partir d'une expérience Monte Carlo peut parfois nécessiter le calcul d'un grand nombre de simulations. Ceci n'est pas toujours réalisable. Dans certains cas, le nombre de simulations nécessaire peut être réduit de manière significative en utilisant certaines techniques de réduction de variance que nous allons présenter dans la suite.

### 1.3.2 Monte Carlo conditionnel

#### 1.3.2.1 Contexte général

On appelle aussi cette technique de réduction de variance : réduction de la dimension. S'il est possible de décomposer le vecteur des variables d'entrée  $x$  en deux sous vecteurs  $y$  et  $z$ , ( $x = (y, z)$ ), l'intégrale (1.3.1) peut s'écrire de la manière suivante :

$$\begin{aligned} I &= \int \int g(y, z) f(y, z) dy dz, \\ &= \int \int g(y, z) f_{Z|Y}(y, z) f_Y(y) dz dy, \\ &= \int \mathbb{E} [g(y, Z) | Y = y] f_Y(y) dy, \end{aligned} \tag{1.3.6}$$

où  $f_{Z|Y}$  représente la densité conditionnelle de  $Z$  sachant  $Y$  :

$$f_{Z|Y}(y, z) = \begin{cases} \frac{f(y, z)}{f_Y(y)} & \text{si } f_Y(y) > 0 \\ 0 & \text{sinon.} \end{cases}$$

Si l'on sait calculer analytiquement la loi marginale  $f_Y(y) = \int f(y, z) dz$  et l'espérance conditionnelle  $\mathbb{E} [g(Y, Z) | Y]$ , l'estimateur de  $I$  est alors :

$$\hat{I}_n^{cond} = \frac{1}{n} \sum_{i=1}^n \mathbb{E} [g(Y_i, Z) | Y_i], \quad Y_i \text{ v.a.i.id} \sim f_Y$$

et sa variance :

$$\text{Var} \left( \hat{I}_n^{cond} \right) = \frac{1}{n} \text{Var} \left( \mathbb{E} [g(Y, Z) | Y] \right)$$

La réduction de la variance par rapport à un Monte Carlo direct est assurée par l'inégalité de Jensen conditionnelle :

$$\begin{aligned} \text{Var} \left( \mathbb{E} [g(Y, Z) | Y] \right) &= \mathbb{E} \left[ \mathbb{E}^2 [g(Y, Z) | Y] \right] - \mathbb{E}^2 \left[ \mathbb{E} [g(Y, Z) | Y] \right] \\ &\leq \mathbb{E} [g^2(Y, Z)] - I^2 \\ \Rightarrow \text{Var} \left( \hat{I}_n^{cond} \right) &\leq \text{Var} \left( \hat{I}_n^{MC} \right). \end{aligned}$$

### 1.3.2.2 Contexte fiabilité des structures

En fiabilité des structures, le principe est de supprimer la fonction indicatrice qui apparaît dans (1.0.1) afin d'obtenir une formulation faisant intervenir la fonction de répartition d'une des variables du modèle. Cette technique nous permet de ramener une intégrale de la forme (1.0.1) à une intégrale de la forme (1.3.1). On cherche à passer de  $P_f = \int \mathbf{1}_{\{G(x) \leq 0\}} f_X(x) dx$  à  $P_f = \int_D g(z) f_Z(z) dz$ ,  $X = (Y, Z)$ . Si la loi marginale  $f_Z$  est connue,  $P_f$  s'écrit :

$$P_f = \int \left( \int \mathbf{1}_{G(y,z) \leq 0} f_{Y|Z}(y, z) f_Z(z) dy \right) dz$$

En posant :

$$g(z) = \int \mathbf{1}_{G(y,z) \leq 0} f_{Y|Z}(y, z) dy, \quad (1.3.7)$$

et si  $g$  définie comme ci-dessus est régulière, nous avons bien une intégrale de type (1.3.1). Ce passage de (1.0.1) à (1.3.1) peut être réalisé sous les deux conditions suffisantes suivantes :

1. la fonction de répartition conditionnelle à une variable ( $x_1$  par exemple) est connue, c'est-à-dire, en posant  $Z = (X_2, \dots, X_d)$ ,  $F_{X_1|Z}(x|Z)$  est connue. En particulier, cette condition est remplie si  $X_1$  est indépendante de  $(X_2, \dots, X_d)$ ;
2. il faut pouvoir isoler une variable dans la fonction de défaillance, la surface de défaillance  $G(x) = G(x_1, z) = 0$  peut alors s'écrire  $x_1 = \tilde{G}(z)$ , et le domaine de défaillance  $G(x) < 0$  est identique au domaine défini par  $x_1 < \tilde{G}(z)$ .

Sous ces conditions, la probabilité de défaillance (1.0.1) devient :

$$P_f = \int F_{X_1|Z}(\tilde{G}(z), z) f_Z(z) dz.$$

L'estimateur de  $P_f$  est alors :

$$\hat{P}_{f_n} = \frac{1}{n} \sum_{i=1}^n F_{X_1|X_2^i, \dots, X_d^i}(\tilde{G}(X_2^i, \dots, X_d^i), X_2^i, \dots, X_d^i),$$

et sa variance :

$$\text{Var}(\hat{P}_{f_n}) = \frac{1}{n} \text{Var}\left(F_{X_1|X_2^i, \dots, X_d^i}(\tilde{G}(X_2^i, \dots, X_d^i), X_2^i, \dots, X_d^i)\right).$$

Remarques :

1. Si  $X_1$  n'est pas indépendante des autres variables du modèle, la fonction de répartition doit utiliser à chaque étape les réalisations conditionnelles des autres variables ;
2. Si la fonction  $\tilde{G}$  n'est pas explicite, il faut trouver le ou les  $x_1^*$  qui satisfont  $G(x_1^*, x_2, \dots, x_d) = 0$ . Cela implique l'utilisation d'algorithmes de recherche de racines, ce qui peut s'avérer coûteux.

La figure (??) illustre le principe de la réduction de la dimension.

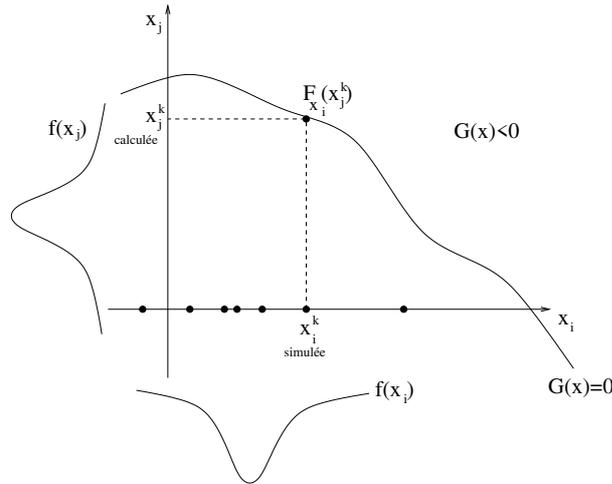


FIG. 1.3.2 – Monte Carlo conditionnel

### 1.3.3 Corrélation de variables

La corrélation entre deux variables aléatoires  $X$  et  $Y$  est donnée par :

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}, \quad (1.3.8)$$

et son estimateur empirique :

$$\hat{\rho}_{XY} = \frac{\sum_{i=1}^n \left( X_i - \frac{1}{n} \sum_{j=1}^n X_j \right) \left( Y_i - \frac{1}{n} \sum_{j=1}^n Y_j \right)}{\left( \sum_{i=1}^n \left( X_i - \frac{1}{n} \sum_{j=1}^n X_j \right)^2 \right)^{1/2} \left( \sum_{i=1}^n \left( Y_i - \frac{1}{n} \sum_{j=1}^n Y_j \right)^2 \right)^{1/2}}$$

#### 1.3.3.1 Variables corrélées

Cette méthode consiste à décomposer la fonction  $g$  en différence de deux fonctions  $g_1$  et  $g_2$  telles que  $g_1(X)$  et  $g_2(X)$  soient fortement positivement corrélés ( $\text{cov}(g_1(X), g_2(X)) > 0$ ).  $I$  s'écrit alors :

$$I = \int g_1(x) f(x) dx - \int g_2(x) f(x) dx.$$

L'estimateur de  $I$  par cette technique s'écrit :

$$\hat{I}_n^{corr} = \frac{1}{n} \sum_{i=1}^n (g_1(X_i) - g_2(X_i)), \quad X_i \sim f, \quad (1.3.9)$$

et sa variance est donnée par :

$$\text{Var} \left( \hat{I}_n^{corr} \right) = \text{Var} (g_1(X)) + \text{Var} (g_2(Y)) - 2\text{cov} (g_1(X), g_2(Y)). \quad (1.3.10)$$

Si  $\text{cov}(g_1(X), g_2(X)) > 0$ , alors  $\text{Var} \left( \hat{I}_n^{corr} \right) \leq \text{Var} \left( \hat{I}_n^{MC} \right)$ .

### 1.3.3.2 Variables antithétiques

Cette méthode vient de Hammersley et Morton [Rub 81], l'idée des variables antithétiques consiste à calculer deux estimations différentes de la quantité d'intérêt et de telle manière que les deux estimations soient corrélées négativement. Leur moyenne empirique sera ensuite substantiellement plus précise que chacune d'elles prise individuellement.

La méthode est décrite en dimension 1. Supposons que  $U$  soit une variable aléatoire uniforme sur  $[0, 1]$  utilisée dans la simulation d'une variable aléatoire  $X$  par la méthode de fonction de répartition inverse, les réalisations de  $U$  seront alors utilisées deux fois,  $U$  et  $1 - U$  suivants toutes deux des lois uniformes sur  $[0, 1]$  (*i.e.*  $X = F_X^{-1}(U)$  et  $X' = F_X^{-1}(1 - U)$ ). L'estimateur de  $I$  s'écrit alors :

$$\hat{I}_n^{AT} = \frac{1}{2n} \sum_{i=1}^n \left( g(X_i) + g(X'_i) \right).$$

La variance de cet estimateur s'écrit :

$$\text{Var} \left( \hat{I}_n^{AT} \right) = \frac{1}{2n} \left[ \text{Var} (g(X)) + \text{Cov} (g(F_X^{-1}(U)), g(F_X^{-1}(1 - U))) \right].$$

On en déduit que lorsque les variables  $g(F_X^{-1}(U))$  et  $g(F_X^{-1}(1 - U))$  sont négativement corrélées :

$$\text{Cov}(g(F_X^{-1}(U)), g(F_X^{-1}(1 - U))) \leq 0,$$

il y a réduction de variance par rapport à une méthode de Monte Carlo directe à  $2n$  simulations :

$$\text{Var} \left( \hat{I}_n^{AT} \right) \leq \text{Var} \left( \hat{I}_{2n} \right).$$

Cette corrélation est bien négative si  $g \in L^1$  et si  $g$  est monotone.

Remarquons que n'importe quelle transformation  $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$  pour laquelle les lois de  $X$  et  $T(X)$  sont identiques convient.

### 1.3.3.3 Variables de contrôle

Cette technique [Liu 01] consiste à employer une variable de contrôle de la forme  $C(X)$  avec  $C$  fonction donnée de telle manière que  $g(X)$  et  $C(X)$  soient corrélées et  $\mathbb{E}[C(X)] = \mu$  connue.

$I$  peut alors s'écrire :

$$I = \mathbb{E}[g(X) - bC(X)] + b\mathbb{E}[C(X)].$$

L'estimateur de  $I$  a la forme suivante :

$$\hat{I}_n^{CV} = \frac{1}{n} \sum_{i=1}^n (g(X_i) - bC(X_i)) + b\mu, \quad X_i \sim f$$

Cet estimateur est sans biais et sa variance est :

$$\text{Var}(\hat{I}_n^{CV}) = \frac{1}{n} (\text{Var}(g(X)) + b^2 \text{Var}(C(X)) - 2b \text{Cov}(g(X), C(X)))$$

L'utilisation d'une variable de contrôle ne garantit pas de réduction de variance par rapport à un Monte Carlo direct. Il faut que  $\text{Var}(g(X) - bC(X)) \leq \text{Var}(g(X))$ ,  $g$  et  $C$  doivent être fortement positivement corrélées, ce qui nécessite une connaissance préalable de  $g$  et un choix approprié de  $C$  (une surface de réponse peut convenir).

Le coefficient de régression optimal est :

$$b = \frac{\text{Cov}(g(X), C(X))}{\text{Var}(C(X))},$$

et dans ce cas,

$$\text{Var}(g(X) - bC(X)) = (1 - \rho_{XC}^2) \text{Var}(g(X)) < \text{Var}(g(X)),$$

où  $\rho_{XC}$  est le coefficient de corrélation entre  $X$  et  $C$  donné par l'équation (1.3.8). On peut utiliser un estimateur de  $b$  (cf. chapitre 5).

### 1.3.4 Stratification

#### 1.3.4.1 Contexte général

L'échantillonnage stratifié [Rub 81] est une autre technique de réduction de variance très utilisée en simulation. L'idée de base est de partitionner le domaine d'intégration en plusieurs sous-domaines disjoints appelés strates. On intègre alors  $g$  sur chacune des strates avec des pondérations fonctions des strates. Cette méthode permet de faire varier le nombre de tirages en fonction des zones de l'espace. Ceci permet, par exemple, de s'assurer que l'on a des tirages situés dans les queues de distribution.

Soit une partition  $\{D_i, 1 \leq i \leq m\}$  de  $\mathbb{R}^d$  telle que :

$$D = \bigcup_{i=1}^m D_i, \quad D_i \cap D_j = \emptyset, \quad i \neq j$$

On définit le poids de la strate  $i$  par  $p_i = P(X \in D_i)$ . les  $p_i$  sont supposés connus.  $I$  s'écrit alors :

$$I = \sum_{i=1}^m p_i \mathbb{E}[g(X)|X \in D_i] = \sum_{i=1}^m p_i I_i.$$

Chaque intégrale  $I_i = \int_{D_i} \frac{g(x)}{p_i} f_X(x) dx$  peut être estimée par une méthode de Monte Carlo directe avec  $n_i$  simulations ( $\sum_{i=1}^m n_i = n$ ). L'estimateur de  $I$  s'écrit alors :

$$\hat{I}_n^{ST} = \sum_{i=1}^m p_i \hat{I}_i$$

cet estimateur est sans biais et convergent si chaque  $n_i$  tend vers  $+\infty$  quand  $n$  tend vers  $+\infty$ . Si les estimateurs  $\hat{I}_i$  sont indépendants, la variance de  $\hat{I}_n^{ST}$  est :

$$\text{Var} \left( \hat{I}_n^{ST} \right) = \sum_{i=1}^m \frac{p_i^2}{n_i} \text{Var} (g(X)|X \in D_i).$$

Le gain en variance par rapport à une méthode de Monte Carlo directe n'est pas assuré par cette méthode. La réduction de variance dépend du choix des strates et du nombre de tirages affectés à chacune d'entre elles.

Le minimum de  $\text{Var} \left( \hat{I}_n^{ST} \right)$  est atteint pour :

$$n_i = n \frac{p_i \text{Var} (g(X)|X \in D_i)^{1/2}}{\sum_{j=1}^m p_j \text{Var} (g(X)|X \in D_j)^{1/2}}, \quad 1 \leq i \leq m$$

et vaut :

$$\hat{I}_n^{ST} = \frac{1}{n} \left( \sum_{i=1}^m p_i \text{Var} (g(X)|X \in D_i)^{1/2} \right)^2.$$

Cependant, le choix optimal des  $n_i$  (pour  $m$  fixe) nécessite de connaître les variances conditionnelles, ce qui n'est pas toujours le cas ; on peut alors les estimer par un Monte Carlo direct. Il faut prendre garde qu'un « mauvais » choix des  $n_i$  peut augmenter la variance de l'estimateur.

Notons que le choix  $n_i = np_i$  diminue toujours la variance. Il consiste à prendre la taille de l'échantillon de la strate  $i$  proportionnelle à sa probabilité.

Un cas particulier est de découper le domaine d'intégration en strates équiprobables et d'effectuer autant de tirages dans toutes les strates. C'est-à-dire, pour la strate  $i$ ,  $p_i = 1/m$  et  $n_i = n/m$ .

#### 1.3.4.2 Contexte fiabilité des structures

En fiabilité des structures, la méthode de stratification s'applique de manière similaire au contexte général avec  $g$  qui est une indicatrice. Néanmoins, la variance de l'estimateur s'écrit avec une forme particulière :

$$\text{Var} \left( \hat{P}_{f_n}^{ST} \right) = \sum_{i=1}^m p_i^2 \frac{P_f^i (1 - P_f^i)}{n_i}, \quad (1.3.11)$$

avec  $P_f^i = \int_{D_i} \frac{\mathbf{1}_{G(x) \leq 0}}{p_i} f_X(x) dx$  Un illustration du principe de simulation par stratification est donnée figure (1.3.3).

#### 1.3.4.3 Hypercube latin

La méthode d'hypercube latin [MBC 79] n'est pas adaptée à l'estimation de faibles probabilités, en revanche, elle peut se révéler très efficace pour réduire la variance quand c'est la

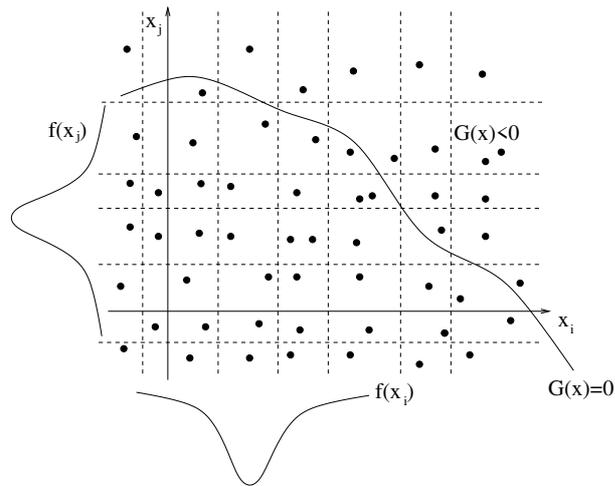


FIG. 1.3.3 – *Monte Carlo stratifié*

partie centrale qui doit être échantillonnée. Cette méthode reprend le principe de la stratification, à ceci près que les tirages ne sont pas effectués dans toutes les strates mais seulement sur un sous ensemble défini selon certains critères (un seul tirage dans chaque ligne et dans chaque colonne). De nombreuses références à ce sujet existent, par exemple [Ste 87], et plus récemment [TH 00], [OSD 03]. La figure (1.3.4) illustre le principe de l'hypercube latin dans un contexte de fiabilité des structures.

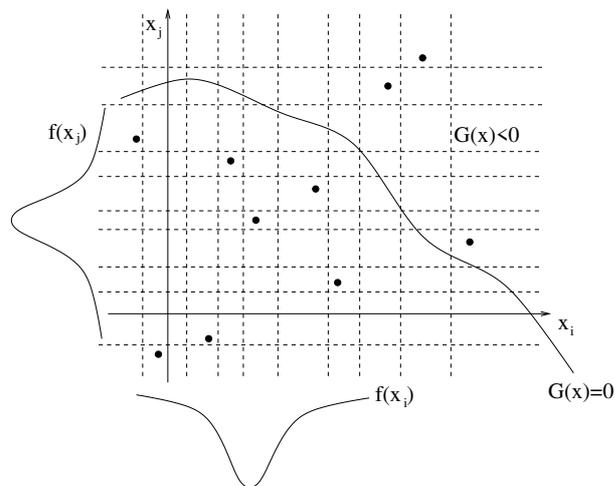


FIG. 1.3.4 – *Monte Carlo Hypercube Latin*

### 1.3.5 Simulation par somme de Riemann

Cette méthode de simulation [Phi 97] permet d'approcher l'intégrale (1.3.1) par :

$$\hat{I}_n^{SR} = \sum_{i=1}^{n-1} (X_{(i+1)} - X_{(i)})g(X_{(i)})f(X_{(i)}), \quad (1.3.12)$$

où  $(X_i, \dots, X_n)$  est un échantillon i.id suivant  $f$  et  $(X_{(1)}, \dots, X_{(n)})$  les statistiques d'ordre de cet échantillon.

La variance de cet estimateur est finie si  $\mathbb{E}[g^2(X)]$  est finie. Si  $g$  est bornée, dérivable et de dérivée bornée, alors l'estimateur est asymptotiquement sans biais et la vitesse de convergence de  $\text{Var}(\hat{I}_n^{SR})$  vers 0 est en  $O(n^{-2})$ .

Les conditions sur  $g$  sont très restrictives et ne permettent pas d'utiliser cette méthode de simulation dans un contexte de fiabilité des structures. De plus, lorsque les conditions sur  $g$  ne sont pas respectées, la convergence de l'estimateur de Riemann vers  $I$  peut être très lente.

### 1.3.6 Tirage d'importance

Le tirage d'importance ou échantillonnage préférentiel est une méthode (ou plutôt une famille de méthodes) de réduction de variance qui préconise de simuler le processus non pas à l'aide des lois des variables réelles (Monte Carlo direct), mais d'utiliser d'autres lois dont le but est de concentrer les tirages dans les régions de l'espace les plus intéressantes pour le calcul de l'intégrale. Il suffit de pondérer les résultats par des rapports de vraisemblance pour supprimer le biais de l'estimateur.

#### 1.3.6.1 Contexte général

On cherche à calculer  $I = \int_D g(x)f(x)dx$ .

Soit  $\tilde{f}$  telle que  $\tilde{f} > 0$  et  $\int_{\mathbb{R}^d} \tilde{f}(x)dx = 1$ ,  $\tilde{f}$  est appelée densité d'importance.  $I$  s'écrit :

$$I = \int_D \frac{g(x)f(x)}{\tilde{f}(x)} \tilde{f}(x)dx.$$

Il vient alors l'égalité suivante :

$$I = \mathbb{E}_f[g(X)] = \mathbb{E}_{\tilde{f}}\left[\frac{g(Y)f(Y)}{\tilde{f}(Y)}\right] \quad \text{avec } X \sim f \text{ et } Y \sim \tilde{f}. \quad (1.3.13)$$

L'estimateur de  $I$  par le tirage d'importance est donné par :

$$\hat{I}_n^{ti} = \frac{1}{n} \sum_{i=1}^n \frac{g(Y_i)f(Y_i)}{\tilde{f}(Y_i)}, \quad Y_i \sim \tilde{f}, \quad (1.3.14)$$

et sa variance :

$$\text{Var} \left( \hat{I}_n^{ti} \right) = \frac{1}{n} (\text{Var}_{\tilde{f}} [g(Y)f(Y)/\tilde{f}(Y)]) = \frac{1}{n} \left( \int \frac{g(x)^2 f(x)^2}{\tilde{f}(x)} dx - I^2 \right) \quad (1.3.15)$$

Remarques :

1. Pour que l'estimateur  $\hat{I}_n^{ti}$  soit sans biais, il faut que le support de la densité d'importance  $\tilde{f}$  contienne le support de la densité initiale  $f$ .
2. L'échantillonnage préférentiel ne garantit pas une réduction de variance par rapport à la simulation simple. Le choix de la densité d'importance influe directement sur la variance de l'estimateur. Il faut établir une comparaison entre  $\text{Var}_{\tilde{f}} [g(X)f(X)/\tilde{f}(X)]$  et  $\text{Var}_f [g(X)]$ .

En théorie, il est possible d'améliorer la précision de l'estimateur. Le minimum de la variance  $\text{Var} \left( \hat{I}_n^{ti} \right)$  est atteint pour la densité d'importance suivante, dite *densité optimale* :

$$f^*(x) = \frac{|g(x)| f(x)}{\int |g(y)| f(y) dy} \quad (1.3.16)$$

Le dénominateur de la densité  $f^*$  est aussi difficile à estimer que  $I$  (c'est  $I$  lorsque  $g$  est positive).

Si  $g(x) \geq 0$ , nous avons  $f^*(x) = \frac{g(x)f(x)}{I}$  et ainsi  $\text{Var} \left( \hat{I}_n^{ti} \right) = 0$ .

Pour le choix de la densité d'importance, il serait intéressant de s'approcher de la densité optimale. L'idée serait d'acquérir des connaissances sur le modèle pour définir la densité d'importance. Il existe de nombreuses stratégies pour construire une densité d'importance. Nous les aborderons dans le chapitre suivant.

### 1.3.6.2 Contexte fiabilité des structures

En fiabilité des structures,  $g$  est de la forme  $\mathbf{1}_{G(x) \leq 0}$ . La densité optimale s'écrit :

$$f^*(x) = \frac{\mathbf{1}_{G(x) \leq 0}(x) f(x)}{P_f} \quad (1.3.17)$$

La densité optimale correspond à la loi conditionnelle de  $X$  sachant que  $X$  est dans le domaine de défaillance  $\{G(X) \leq 0\}$ . L'idée est d'utiliser des nouvelles lois favorisant les tirages dans le domaine de défaillance. Il faudrait pouvoir pousser vers les défaillances, mais sans trop forcer car dans ce cas la variance de l'estimateur deviendrait trop importante. Une illustration du principe de simulation par tirage d'importance est donnée figure (1.3.5).

L'application du tirage d'importance en fiabilité des structures a fait l'objet de nombreuses stratégies de construction de la densité d'importance ; nous les verrons dans le chapitre suivant.

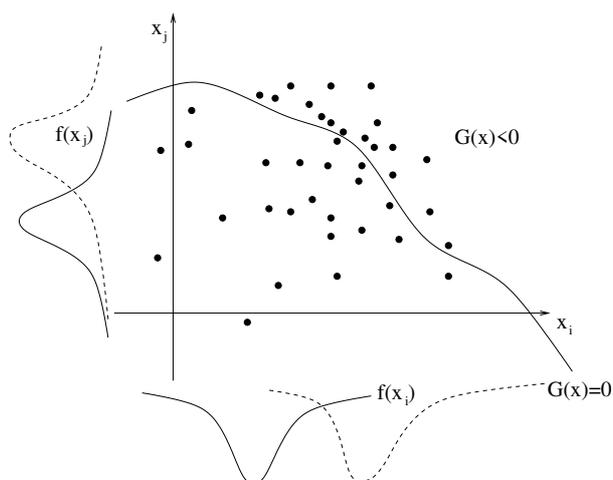


FIG. 1.3.5 – Monte Carlo Tirage d'importance, densités originales en pleins, densités d'importance en pointillés

### 1.3.6.3 Découpage du domaine d'intégration

Cette méthode est en fait une généralisation du tirage d'importance; elle est peu utilisée en pratique car elle implique que l'on sache calculer  $I$  sur une partie du domaine d'intégration.

$$I = \int_{D_1} g(x)f(x)dx + \int_{D_2} g(x)f(x)dx, \quad D = D_1 \cup D_2, \quad D_1 \cap D_2 = \emptyset.$$

On suppose que l'on sait calculer  $I_1 = \int_{D_1} g(x)f(x)dx$ . On introduit une nouvelle densité  $\tilde{f}$  telle que  $\int_{D_1} \tilde{f}(x)dx = p$ . L'intégrale devient :

$$I = I_1 + (1 - p)\mathbb{E} \left[ \frac{g(X)f(x)}{\tilde{f}(x)} \right], \quad X \sim \tilde{f}, \quad (1.3.18)$$

L'estimateur de  $I$  s'écrit :

$$\hat{I}_n^{dec} = I_1 + (1 - p)\frac{1}{n} \sum_{i=1}^n \frac{g(X_i)f(X_i)}{\tilde{f}(X_i)}, \quad X_i \sim \tilde{f}, \quad (1.3.19)$$

et sa variance :

$$\text{Var} \left( \hat{I}_n^{dec} \right) = \frac{1}{n}(1 - p)^2 \left[ \int_{D_2} \frac{g^2(x)f^2(x)}{\tilde{f}(x)} dx - \left( \int_{D_2} g(x)f(x)dx \right)^2 \right]. \quad (1.3.20)$$

On a la relation suivante :

$$\text{Var} \left( \hat{I}_n^{dec} \right) \leq (1 - p)^2 \text{Var} \left( \hat{I}_n^{ti} \right). \quad (1.3.21)$$

### 1.3.7 Simulation directionnelle

Cette méthode s'applique au calcul de (1.0.1) et non de (1.3.1). On ne peut l'utiliser que dans un cas de fiabilité des structures.

#### Principe

La simulation directionnelle [Bje 88, Mel 94] est une technique de réduction de variance pour les problèmes de fiabilité des structures; Elle s'effectue dans l'espace gaussien (cf. §1.2.2.1), on dispose alors d'un vecteur aléatoire  $U = (U_1, \dots, U_d)$  dont les  $d$  composantes sont gaussiennes indépendantes centrées et réduites; on cherche à calculer la probabilité de défaillance qui s'écrit comme suit :  $P_f = P(H(U) \leq 0)$ .

Cette méthode consiste à générer des directions uniformes et indépendantes issues de l'origine de l'espace gaussien et à calculer l'intégrale conditionnellement à ces directions. Soit  $\Omega_d$  la sphère unité de dimension  $d$  centrée à l'origine, le vecteur  $U$  peut s'écrire  $U = RA$  ( $R \geq 0$ ), où  $R^2$  est une variable aléatoire suivant une loi du  $\chi^2$  à  $d$  degrés de liberté et  $A$  un vecteur aléatoire unitaire indépendant de  $R$  et uniformément distribué sur  $\Omega_d$ .

La probabilité de défaillance se met sous la forme :

$$P_f = P(H(RA) \leq 0) = \int_{\Omega_d} P(H(Ra) \leq 0) f_A(a) da$$

où  $f_A$  représente la densité uniforme sur  $\Omega_d$ .

L'estimateur de la probabilité s'écrit :

$$\hat{P}_f^{dir} = \frac{1}{n} \sum_{i=1}^n P(H(Ra_i) \leq 0) \quad (1.3.22)$$

où  $a_i$  correspondent à des directions uniformes et indépendantes sur  $\Omega_d$ .

La variance de l'estimateur s'exprime par :

$$\text{Var}(\hat{P}_f^{dir}) = \frac{1}{n} \left( \mathbb{E} \left[ (P(H(RA) \leq 0))^2 \right] - P_f^2 \right) \quad (1.3.23)$$

#### Calcul de $P(H(\mathbf{R}a_i) \leq 0)$ :

En principe, on ne connaît pas  $H$  de manière explicite, il est donc nécessaire de repasser par la forme  $G(T(\cdot))$  afin de trouver les valeurs de  $R$  qui annulent la fonction  $H(Ra_i)$ . Si  $r$  est racine unique de l'équation  $H(Ra_i) = 0$ , alors :

$$P(H(Ra_i) \leq 0) = 1 - \chi_d^2(r^2), \text{ si } (G(T(0)) \geq 0)$$

S'il y a plusieurs racines à l'équation, la probabilité s'exprime par :

$$P(H(Ra_i) \leq 0) = \sum_i [(-1)^{i+1} (1 - \chi_d^2(r_i^2))] , \text{ si } (G(T(0)) \geq 0)$$

Cette méthode est d'autant plus efficace que la proportion de la surface  $\Omega_d$  qui génère des directions interceptant la surface d'état limite est grande (surface de défaillance dans l'espace gaussien presque sphérique). En revanche, l'algorithme de recherche de racines peut conduire à un surcoût non négligeable en terme d'évaluations de la fonction  $G$ , il est nécessaire de trouver un compromis entre précision des racines et nombre d'évaluation de  $G$ . On peut retrouver dans la littérature des améliorations de cette méthode et son utilisation en fiabilité des systèmes, dans [NE 00] par exemple.

La figure (1.3.6) illustre le principe de la simulation directionnelle.

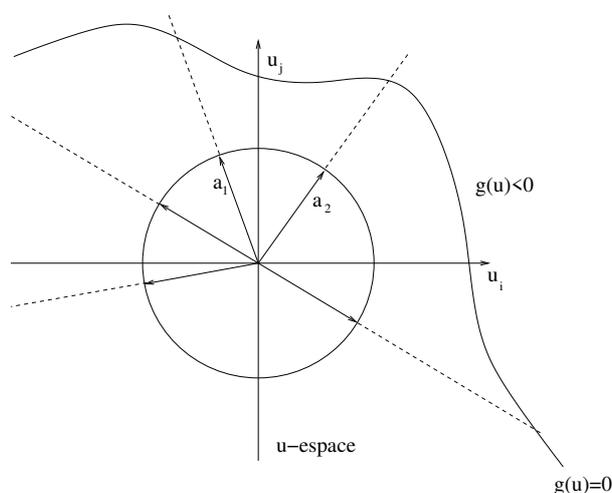


FIG. 1.3.6 – *Simulation directionnelle*

### 1.3.8 Méthodes de quasi Monte Carlo

L'idée des méthodes quasi-Monte Carlo est de remplacer les suites aléatoires des méthodes de Monte Carlo par des suites quasi-aléatoires qui sont des suites déterministes, également appelées suites à discrédance faible. Ces suites ont une meilleure répartition uniforme, c'est-à-dire une discrédance plus faible que les suites aléatoires. Ces méthodes ne sont pas adaptées aux calculs des faibles probabilités car elles permettent d'obtenir des tirages équirépartis et pas forcément de s'approcher des queues de distribution. De plus, elles ne permettent pas d'obtenir une variance pour l'estimateur (les tirages ne sont plus considérés comme indépendants), ni d'intervalles de confiance ; on peut juste avoir une majoration de l'erreur de l'estimation d'une intégrale grâce à l'inégalité de Koksma-Hlwaka.

Nous ne donnerons ici que la définition de la discrédance ainsi que quelques exemples de suites à discrédance faible. La discrédance d'une suite mesure la déviation de cette suite par rapport à la distribution uniforme, en d'autres termes, elle mesure l'irrégularité de la

distribution de la suite ; il existe plusieurs type de discr pance [Tuf 97], nous ne nous int ressons qu'  la discr pance   l'origine.

**D finition 1.3.1** *Discr pance   l'origine*

Soit  $\xi = (\xi_n)_{n \in \mathbb{N}}$  une suite   valeurs dans  $[0, 1]^s$ . La discr pance des  $N$  premiers termes de  $\xi$  est alors d finie par :

$$D_N^*(\xi) = \sup_{x \in [0,1]^d} \left| \hat{F}_N(x) - F(x) \right|,$$

o   $F$  est la fonction de r partition de la loi uniforme sur  $[0, 1]^d$ ,  $F(x) = \prod_{i=1}^d x_i \forall x \in [0, 1]^d$  ; et  $\hat{F}_N$  est la fonction de r partition empirique de la suite  $(\xi_1, \dots, \xi_N)$  d finie par  $\hat{F}_N(x) = \prod_{i=1}^d \frac{\text{card}\{j, \xi_{j,i} \leq x_i\}}{N}$ .

**Proposition 1.3.1** *In galit  de Koksma-Hlwaka*

Si  $g$  est une fonction   variation finie au sens de Hardy et Krausse [Nie 92] de variation  $V(g)$ , alors :

$$\left| \frac{1}{n} \sum_{i=1}^n g(\xi_i) - \int_{[0,1]^d} g(x) dx \right| \leq V(g) D_N^*(\xi), \forall n \geq 1,$$

o   $D_N^*(\xi)$  est la discr pance   l'origine de  $\xi$ .

Voici deux exemples classiques de suites   discr pance faible ; mais parmi les plus utilis es dans la litt rature [Nie 92], on peut citer les suites de Van Der Corput, les suites de Halton, les suites de Faure, les suites de Sobol, les suites de Hammersley, les suites SQRT, les suites de Niederreiter, les suites de Lapeyre-Pag s.

**Suite de Van Der Corput et suite de Halton**

Soit  $n \in \mathbb{N}$  et son d veloppement en base  $b$  :

$$n = \sum_{j=1}^{\infty} a_j(n) b^j.$$

Soit la fonction radicale inverse  $\phi_b$  en base  $b$  d finie par :

$$\phi_b(n) = \sum_{j=1}^{\infty} a_j(n) b^{-j-1} \in [0, 1]$$

La suite de Van Der Corput est donn e par :

$$\xi_n = \phi_2(n), \quad n \geq 0$$

La suite de Halton est une g n ralisation de la suite de Van Der Corput. Pour  $d \geq 1$ , on consid re  $b_1, \dots, b_s$  les entiers premiers. La suite de Halton dans la base  $b_1, \dots, b_d$  est la suite infinie  $\xi_1, \xi_2, \dots$  avec

$$\xi_n = (\phi_{b_1}, \dots, \phi_{b_d}) \in [0, 1]^d.$$

La discrédance à l'origine de la suite de Halton est en  $O(N^{-1}(\log(N))^d)$ .

### Suite SQRT ou suite $\{n \cdot \alpha\}$

Soit  $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{R}^d$  tel que  $1, \alpha_1, \dots, \alpha_d$  linéairement indépendants sur  $\mathbb{Q}$ . La suite  $n \cdot \alpha$  est donnée par :

$$\xi_n = \{\{n \cdot \alpha_1\}, \dots, \{n \cdot \alpha_d\}\}$$

où  $\{n \cdot \alpha\} = n \cdot \alpha \pmod{1}$ . On peut choisir par exemple  $\alpha = (\sqrt{p_1}, \dots, \sqrt{p_d})$  avec  $p_i, 1 \leq i \leq d$  les  $d$  premiers nombres premiers.

Pour tout  $\epsilon > 0$ , la discrédance d'une suite  $\{n \cdot \alpha\}$  est en  $O(N^{-1}(\log(N))^{1+d+\epsilon})$  pour presque tout  $\alpha \in \mathbb{R}^d$ . Aucun  $\alpha$  fournissant une discrédance  $O(N^{-1}(\log(N))^{d+1})$  n'est connu pour  $d \geq 2$ . Si les  $\alpha_i$  sont algébriques, alors la discrédance est en  $O(N^{-1+\epsilon})$  pour  $\epsilon > 0$ .

## 1.3.9 Combinaisons de méthodes

### 1.3.9.1 Avec la méthode de Monte Carlo conditionnel

En fiabilité des structures, quand on cherche à calculer une intégrale de type (1.0.1) et que l'on peut appliquer une méthode de Monte Carlo conditionnel présentée en 1.3.2, on se ramène en fait à une intégrale de type (1.3.1), et on a alors la possibilité d'appliquer toutes les techniques de réduction de variance. Ainsi on peut combiner la réduction de la dimension avec des méthodes de stratification, mais aussi avec le tirage d'importance.

### 1.3.9.2 Méthodes d'approximations et de simulations

Les méthodes FORM/SORM peuvent conduire à de bonnes approximations de la probabilité de défaillance. Elles apportent ainsi des informations dont on pourrait se servir pour réaliser des simulations de Monte Carlo. Par exemple, on peut combiner le tirage d'importance avec les résultats de FORM, à savoir les coordonnées du ou des points de conception. La densité d'importance sera alors centrée autour du point de conception. Nous reviendrons sur ces combinaisons dans le chapitre suivant.

### 1.3.9.3 Tirage directionnel et tirage d'importance

Dans [Bje 89, Mel 90, DMG 90], il est proposé de combiner la simulation directionnelle avec le tirage d'importance, ces deux méthodes de réduction de variance sont respectivement présentées en 1.3.7 et 1.3.6. L'idée est de générer les directions  $a$  suivant une densité d'importance  $\tilde{f}_A(a)$  sur la sphère unité (au lieu de les tirer uniformément comme dans la simulation directionnelle classique),  $f_A$  représente la densité uniforme sur la sphère unité  $\Omega_d$  en dimension

*d.* L'objectif est de générer les tirages dans les directions qui contribuent le plus à la défaillance. La probabilité s'écrit :

$$P_f = P(H(RA) \leq 0) = \int_{\Omega_d} P(H(Ra) \leq 0) \frac{f_A(a)}{\tilde{f}_A(a)} \tilde{f}_A(a) da.$$

L'estimateur s'écrit :

$$\hat{P}_f^{dir-ti} = \frac{1}{n} \sum_{i=1}^n P(H(RA_i) \leq 0 | A_i) \frac{f_A(A_i)}{\tilde{f}_A(A_i)}, \quad A_i \sim \tilde{f}_A, \quad (1.3.24)$$

et sa variance est :

$$\text{Var} \left( \hat{P}_f^{dir-ti} \right) = \frac{1}{n} \left( \mathbb{E}_{\tilde{f}_A} \left[ (P(H(RA) \leq 0 | A))^2 \left( \frac{f_A(A_i)}{\tilde{f}_A(A_i)} \right)^2 \right] - P_f^2 \right). \quad (1.3.25)$$

De la même manière que pour le tirage d'importance classique, le choix de la densité d'importance peut s'avérer délicat et n'assure pas systématiquement une réduction de variance.

#### 1.3.9.4 Simulation axis orthogonale

C'est une méthode de simulation qui combine simulation directionnelle et les résultats des calculs FORM, à savoir les coordonnées du point de conception, l'indice de Hasofer-Lind et les calculs du gradient de la fonction d'état limite au point de conception. La méthode consiste tout d'abord à se placer dans l'espace gaussien où le point de conception  $u^*$  a pour coordonnées  $(0, 0, \dots, \beta_{HL})$  (transformation de l'espace d'entrée des variables décrites en 1.2.2.1 et rotation du repère).

L'intégrale (1.0.1) s'écrit :

$$P_f = \int_{H_{axis}(u) \leq 0} \varphi_d(u) du,$$

où  $\varphi_d$  est la densité multinormale centrée réduite de dimension  $d$ , et  $H_{axis}$  la fonction d'état limite dans ce nouveau repère :  $H_{axis}(u) = G(T^{-1}(\Omega u))$  où  $\Omega$  est une matrice de rotation.

Les simulations sont réalisées conditionnellement à des directions correspondant aux  $d - 1$  coordonnées dans le nouveau repère. En posant  $\tilde{U} = (U_1, \dots, U_{d-1})$ , la probabilité de défaillance s'écrit :

$$P_f = \int P \left( H_{axis}(\tilde{U}, U_d) \leq 0 | \tilde{U} = \tilde{u} \right) \varphi_{d-1}(\tilde{u}) d\tilde{u}. \quad (1.3.26)$$

Le terme  $P \left( H_{axis}(\tilde{U}, U_d) \leq 0 | \tilde{U} = \tilde{u} \right)$  se calcule par recherche des racines de  $H_{axis}(\tilde{u}, u_d) =$

0, il se réécrit :

$$\begin{aligned}
 P\left(H_{axis}(\tilde{U}, U_d) \leq 0 \mid \tilde{U} = \tilde{u}\right) &= P(H_{axis}(\tilde{u}, U_d) \leq 0) \\
 &= \int \mathbf{1}_{\{H_{axis}(\tilde{u}, u_d) \leq 0\}}(\tilde{u}, u_d) \varphi_1(u_d) du_d. \quad (1.3.27)
 \end{aligned}$$

### 1.3.10 Organigramme des méthodes

La figure (1.3.7) permet de visualiser les méthodes de calculs d'une probabilité de défaillance et les liens qui existent entre-elles.

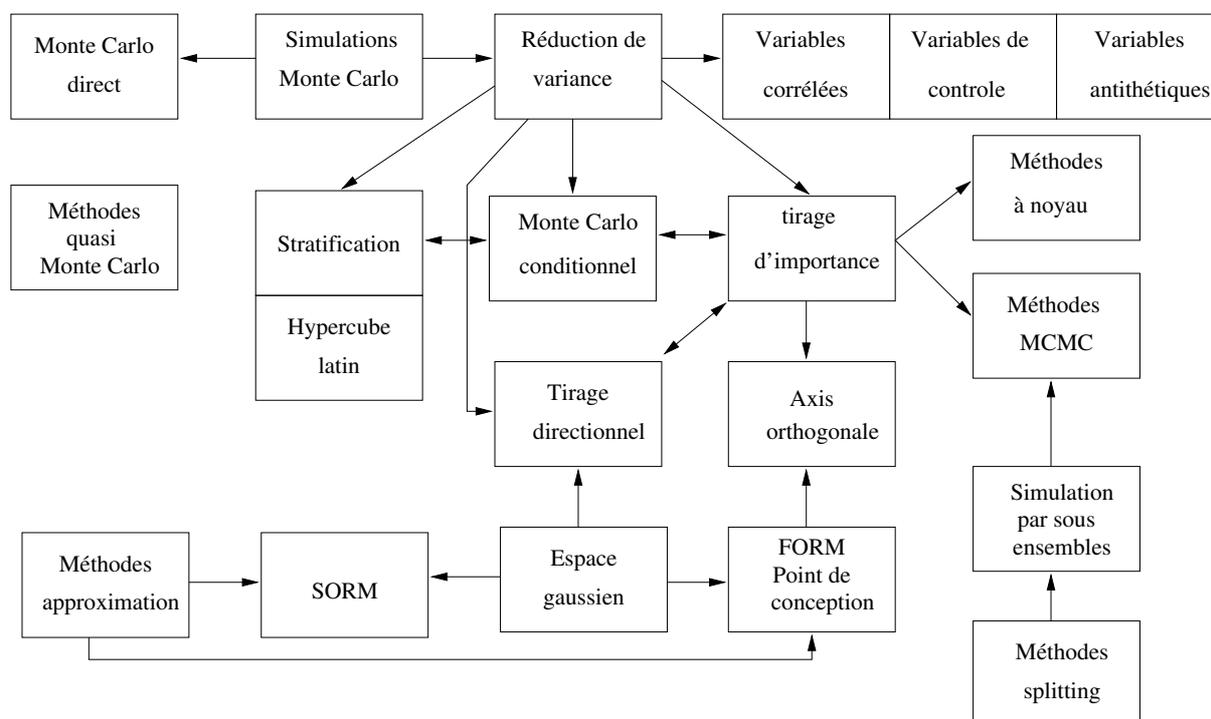


FIG. 1.3.7 – Organigramme des méthodes en fiabilité des structures (d'après [GG 99])

Les quatre méthodes de droite seront décrites dans les chapitres 2 et 3.

## 1.4 Synthèse partielle

Dans ce chapitre, de nombreuses méthodes utilisées en fiabilité des structures ont été présentées. Les principales observations que l'on peut en extraire sont les suivantes :

- Les méthodes d'approximation ne nécessitent pas un nombre élevé de simulations pour estimer une faible probabilité de rupture. En revanche, elles ne permettent pas de contrôler l'erreur commise sur le résultat. Cette erreur peut être grande lorsque l'on n'approxime pas correctement la surface d'état limite (surface fortement non

linéaire par exemple) et/ou lorsqu'il existe plusieurs points de conception (algorithme d'optimisation globale coûteux), ou bien lorsque l'on omet un ou plusieurs points de conception. Ainsi, il est préférable de valider et/ou de corriger le résultat de ces méthodes par un ajout de simulations.

- Les méthodes de Monte Carlo peuvent être très gourmande en nombre de simulations et aussi en temps de calcul. Certaines techniques de réduction de variance garantissent un gain en nombre de simulations (par exemple Monte Carlo conditionnelle), mais d'autres ne le garantissent pas toujours (stratification, tirage d'importance); seul un choix réfléchi des paramètres régissant ces techniques (le choix des strates, la densité d'importance) peut permettre une utilisation à bon escient de ces méthodes. C'est le choix et/ou la construction de la densité d'importance qui a motivé le chapitre suivant.
- Pour comparer l'efficacité des méthodes, il est nécessaire d'obtenir des informations sur les variances des estimateurs de la probabilité de défaillance. La mise au point d'estimateurs de la variance des estimateurs est donc un point important. Ils permettent de comparer, à nombre de simulations fixes, les méthodes de simulations de Monte Carlo entre elles mais pas les méthodes de simulations avec les méthodes d'approximation.



## CHAPITRE 2

### TIRAGE D'IMPORTANCE

Parmi les méthodes présentées dans le chapitre précédent, le tirage d'importance (§ 1.3.6) a particulièrement retenu notre attention, cette méthode peut offrir de bons résultats (compromis entre le nombre de simulations et la précision) à condition de bien choisir la densité d'importance  $\tilde{f}$ . Nous commencerons par faire des rappels sur le tirage d'importance, puis nous aborderons les différentes pistes bibliographiques concernant le choix ou la construction de cette densité d'importance dans un contexte général d'estimation d'une espérance mathématique et dans un contexte de fiabilité des structures. Nous détaillerons ensuite particulièrement la construction d'une densité d'importance de manière adaptative.

#### 2.1 Généralités sur le tirage d'importance

##### 2.1.1 Résultats classiques

Nous nous plaçons dans le cas où nous cherchons à estimer une espérance mathématique  $I = \int_{D \subset \mathbb{R}^d} g(x)f(x)dx = \mathbb{E}_f[g(X)]$ . Le principe du tirage d'importance est de générer  $X$ , non pas suivant  $f$  mais suivant une densité d'importance  $\tilde{f}$  et de neutraliser le biais en pondérant les observations par un rapport de vraisemblance.

##### . Estimateur

Si  $g(X) \in L^1_{\tilde{f}}(\mathbb{R}^d)$ , l'estimateur de  $I$  par le tirage d'importance s'écrit :

$$\hat{I} = \frac{1}{n} \sum_{i=1}^n g(X_i) \frac{f(X_i)}{\tilde{f}(X_i)}, \quad X_i \sim \tilde{f}, \quad (2.1.1)$$

$\hat{I}$  converge presque sûrement vers  $I$  quand  $n \rightarrow +\infty$  (loi forte des grands nombres).

Rappelons que pour que l'estimateur  $\hat{I}$  soit sans biais, il faut que le support de la densité d'importance contienne le support de la densité initiale ( $\text{supp}(f) \subset \text{supp}(\tilde{f})$ ).

##### . Variance

La variance de l'estimateur  $\hat{I}$  est donnée par :

$$\text{Var}(\hat{I}) = \frac{1}{n} \text{Var}_{\tilde{f}} \left( g(X) \frac{f(X)}{\tilde{f}(X)} \right) = \frac{1}{n} \left( \int \frac{g(x)^2 f(x)^2}{\tilde{f}(x)} dx - I^2 \right). \quad (2.1.2)$$

Si  $\text{Var}_{\tilde{f}}\left(g(X)\frac{f(X)}{\tilde{f}(X)}\right) = \sigma^2 < +\infty$ , nous pouvons appliquer le théorème central limite et nous obtenons :

$$\sqrt{n}(\hat{I} - I) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2), \quad n \rightarrow +\infty. \quad (2.1.3)$$

Un intervalle de confiance de niveau  $1 - 2\alpha$  pour  $\hat{I}$  est donné par :

$$\left[ \hat{I} - \phi^{-1}(\alpha) \sqrt{\frac{\widehat{\text{Var}}(Z)}{n}}, \hat{I} + \phi^{-1}(\alpha) \sqrt{\frac{\widehat{\text{Var}}(Z)}{n}} \right], \quad (2.1.4)$$

où  $Z = \frac{g(X)f(X)}{\tilde{f}(X)}$  et  $\widehat{\text{Var}}(Z) = \frac{1}{(n-1)} \sum_{i=1}^n \left( g(X_i) \frac{f(X_i)}{\tilde{f}(X_i)} - \hat{I} \right)^2$ ,  $X_i \sim \tilde{f}$ .

### . Densité d'importance optimale

Il existe une densité  $\tilde{f}$ , appelée densité d'importance optimale, proportionnelle à  $|g(x)|f(x)$ , qui minimise la variance de  $\hat{I}$ . Elle s'exprime par :

$$f^*(x) = \frac{|g(x)|f(x)}{\int |g(y)|f(y)dy}. \quad (2.1.5)$$

*Démonstration :*

Le problème est de minimiser  $\frac{1}{n} \left( \int \frac{g(x)^2 f(x)^2}{\tilde{f}(x)} dx - I^2 \right)$  sous la contrainte  $\int \tilde{f}(x) dx = 1$ ; cela revient à la minimisation de l'expression  $\int \frac{g(x)^2 f(x)^2}{\tilde{f}(x)} dx$  sous la même contrainte.

Le Lagrangien de ce problème d'optimisation est :

$$L(\lambda, x) = \int \frac{g(x)^2 f(x)^2}{\tilde{f}(x)} - \lambda \left( \int \tilde{f}(x) dx - 1 \right) \quad \text{où } \lambda \text{ est une constante.}$$

D'après le théorème des multiplicateurs de Lagrange, on obtient l'égalité :

$$\frac{g(x)^2 f(x)^2}{\tilde{f}(x)} - \lambda \tilde{f}(x) = 0.$$

D'où le résultat

$$\lambda = \frac{g(x)^2 f(x)^2}{\tilde{f}^2(x)}. \quad (2.1.6)$$

Ainsi, la fonction  $f^*$  qui minimise la variance vérifie  $f^{*2}(x) = \left( \frac{g(x)^2 f(x)^2}{\lambda} \right)$ .

Comme  $\lambda > 0$  (d'après 2.1.6) et  $\tilde{f} > 0$  (par hypothèse), on obtient :

$$\tilde{f}(x) = \frac{f(x)|g(x)|}{\sqrt{\lambda}}$$

D'après notre contrainte  $\int \tilde{f}(x) dx = 1$ , on a  $\int \frac{f(y)|g(y)|}{\sqrt{\lambda}} dy = 1$  soit  $\sqrt{\lambda} = \int f(y)|g(y)| dy$ .

Finalement, on trouve que :

$$\tilde{f}^*(x) = \frac{f(x)|g(x)|}{\int f(y)|g(y)| dy}$$

Quand  $g$  est positive, on a bien  $\tilde{f}^*(x) = \frac{f(x)g(x)}{I}$  et ainsi :

$$\text{Var}_{\tilde{f}} \left[ \frac{g(X)f(X)}{f^*(X)} \right] = \text{Var}_{\tilde{f}} \left[ \frac{g(X)f(X)}{g(X)f(X)} I \right] = 0, \text{ comme variance d'une constante.} \quad \blacksquare$$

### • Fiabilité des structures

Dans un contexte de fiabilité des structures, la densité optimale est :

$$f^*(x) = \frac{\mathbf{1}_{\{G(x) \leq 0\}} f(x)}{P_f} = f(x|D). \quad (2.1.7)$$

Cette densité est inutilisable en pratique car il faudrait savoir générer des points, suivant la densité initiale, tombant uniquement dans le domaine de défaillance et sans avoir recours à l'évaluation de la fonction d'état limite. De plus, il faudrait connaître a priori la constante de normalisation  $P_f$  qui est précisément ce que l'on cherche à estimer.

#### 2.1.2 Variantes de l'estimateur

Pour estimer  $I$ , il est possible d'utiliser d'autres estimateurs.

##### 2.1.2.1 Première variante

Cet estimateur est obtenu en normalisant le rapport de vraisemblance :

$$\begin{aligned} \hat{I}' &= \frac{\frac{1}{n} \sum_{i=1}^n g(X_i)w(X_i)}{\frac{1}{n} \sum_{i=1}^n w(X_i)} = \frac{\hat{I}}{\bar{w}}, X_i \sim \tilde{f}, \\ &= \frac{1}{\sum_{i=1}^n w(X_i)} \sum_{i=1}^n g(X_i)w(X_i), X_i \sim \tilde{f}, \end{aligned} \quad (2.1.8)$$

où  $w(x) = f(x)/\tilde{f}(x)$  est le rapport de vraisemblance.

Cet estimateur est biaisé et le terme en  $1/n$  du développement de ce biais vaut :

$$\mathbb{E}_{\tilde{f}} [\hat{I}' - I] \approx -\frac{\mathbb{E}_{\tilde{f}} [w(X)(g(X)w(X) - Iw(X))]}{n}. \quad (2.1.9)$$

En effet :

$$\begin{aligned} \mathbb{E} [\hat{I}'] &= \mathbb{E}_{\tilde{f}} \left[ \hat{I} (1 - (\bar{w} - 1 + (\bar{w} - 1)^2) + \dots) \right] \\ &\approx I - \frac{\mathbb{E}_{\tilde{f}} [g(X)w(X)^2 - I]}{n} + I \frac{\mathbb{E}_{\tilde{f}} [w(X)^2 - 1]}{n}. \end{aligned}$$

La variance de  $\hat{I}'$  s'exprime par :

$$\begin{aligned} \text{Var} (\hat{I}') &= \frac{1}{n} \text{Var}_{\tilde{f}} ((g(X) - I)w(X)) \\ &= \frac{1}{n} \int \frac{f(x)(g(x) - I)^2}{\tilde{f}(x)} f(x) dx. \end{aligned} \quad (2.1.10)$$

Un intervalle de confiance de niveau  $1 - 2\alpha$  pour  $\hat{I}'$  est donné par :

$$\left[ \hat{I}' - \phi^{-1}(\alpha) \sqrt{\frac{\widehat{\text{Var}}(V)}{n}}, \hat{I}' + \phi^{-1}(\alpha) \sqrt{\frac{\widehat{\text{Var}}(V)}{n}} \right], \quad (2.1.11)$$

où  $V = (g(X) - I)w(X)$  et  $\widehat{\text{Var}}(V) = \frac{1}{n-1} \sum_{i=1}^n \left( (g(X_i) - \hat{I}')w(X_i) \right)^2$ .

### 2.1.2.2 Deuxième variante

Il est aussi possible d'utiliser l'estimateur suivant :

$$\hat{I}'' = \sum_{i=1}^n g(X_i)w(X_i) - \hat{\beta} \left( \sum_{i=1}^n w(X_i) - 1 \right), \quad X_i \sim \tilde{f}, \quad (2.1.12)$$

où  $\hat{\beta}$  est la pente de la régression linéaire de  $gw$  par rapport à  $w$ . Cet estimateur est aussi biaisé. Nous pourrions nous référer à [Hes 88] pour informations supplémentaires.

### 2.1.3 Autres utilisations du tirage d'importance

Nous avons vu que le tirage d'importance peut servir en tant que méthode de réduction de variance dans l'estimation d'intégrale, et en particulier à l'estimation d'espérance mathématique. Mais il est possible d'utiliser cette méthode pour d'autres problèmes :

1. On souhaite calculer une espérance mathématique mais on ne sait pas générer un échantillon avec la loi souhaitée. C'est souvent le cas en estimation bayésienne, on cherche à estimer une espérance mathématique dont les variables d'entrées doivent être simulées suivant la loi a posteriori des paramètres. Les méthodes de Monte Carlo par chaînes de Markov peuvent répondre à cette attente mais leur convergence peut être assez lente. Le tirage d'importance peut être une alternative intéressante, mais utiliser une densité d'importance arbitraire ne résoudrait pas forcément le problème de convergence lente. Des algorithmes de rééchantillonnage basés sur le tirage d'importance permettent de construire un échantillon : l'algorithme SIR (Sampling Importance Resampling), qui permet de construire un échantillon suivant la densité inconnue et l'algorithme PMC (Population Monte Carlo) [CGP 03] qui est en fait un schéma SIR adaptatif.
2. On souhaite tester plusieurs types de densité initiale. Soit  $f(x, \alpha)$  la densité initiale, paramétrée par  $\alpha$ , et  $\mathbb{E}[h(X, \alpha)]$  la quantité cherchée. L'emploi du tirage d'importance nous permet de générer un seul échantillon suivant la densité d'importance pour mesurer l'effet des différentes densités initiales sur  $\mathbb{E}[h(X, \alpha)]$ . Cet emploi particulier du tirage d'importance demande certaines précautions (cf. § 2.2.2.1).

## 2.2 État de l'art

Nous présentons ici un état de l'art sur la construction d'une densité d'importance. On parle de tirage d'importance adaptatif quand la densité d'importance se construit par étapes. Il existe deux manières de procéder, l'une paramétrique, on choisit une famille de densités, paramétrée par un paramètre  $\gamma$  et on essaye de s'approcher au mieux de la densité d'importance optimale en adaptant ses paramètres ; l'autre non paramétrique basée sur l'utilisation de noyaux régularisants. Dans un contexte de fiabilité des structures, les méthodes adaptatives sont basées sur l'idée que la connaissance du domaine de défaillance augmente avec le nombre de simulations.

### 2.2.1 Tirage d'importance pour la fiabilité des structures

L'utilisation du tirage d'importance dans un contexte de fiabilité des structures a conduit à de nombreuses publications sur la question cruciale : comment exploiter au mieux les informations fournies par les méthodes d'approximation dans la construction de la densité d'importance ? Nous avons déjà énoncé ce principe dans le chapitre précédent (§ 1.3.9.2). Comme dans la simulation axis orthogonale (§ 1.3.9.4), il est souvent proposé de choisir la densité d'importance centrée autour du point de conception. Dans la suite de cette section, nous allons aborder de nombreuses techniques de tirage d'importance utilisant ce point, fournie par les méthodes d'approximation. Nous aborderons aussi le cas de la fiabilité des systèmes.

#### 2.2.1.1 Densité d'importance centrée autour du point de conception

Ces méthodes sont non adaptatives, en revanche elles sont toutes paramétriques. Nous supposons connaître les coordonnées du point de conception dans l'espace d'entrée des variables et dans l'espace gaussien. Tout d'abord, A. Harbitz [Har 83] émet l'idée de centrer la densité d'importance autour du point de conception. Nous présentons les choix de différents auteurs concernant le choix de la forme de la densité d'importance. il est spécifié quand celle-ci n'est pas centrée au point de conception.

Dans [Shi 83], M. Shinozuka propose de choisir une densité uniforme sur une région importante déterminée au préalable, c'est un domaine « rectangulaire » qui couvre la région du maximum vraisemblance autour du point de conception. Cette technique induit un biais dans l'estimateur de la probabilité de défaillance ; en effet, pour que l'estimateur de la probabilité soit sans biais, il faut que le support de la densité d'importance contienne le support de la densité initiale (cf. § 1.3.6.1). Dans [SS 87], il est préconisé de choisir une densité multinormale avec la même matrice de covariance que la densité initiale. Dans [FR 88], les auteurs se

placent dans l'espace gaussien et une densité multinormale réduite est utilisée. R. E. Melchers [Mel 89] propose d'utiliser une densité multinormale avec une matrice de covariance diagonale dont les composantes ne sont pas inférieures à celle de la matrice de covariance de la densité initiale. Y. Ibrahim [Ibr 91] propose d'utiliser la même forme que celle de la densité initiale.

### 2.2.1.2 Tirage d'importance conditionné

Cette variante du tirage d'importance utilise le principe des tirages conditionnés [Har 86, BF 87], qui propose d'exclure les tirages de l'hyper-sphère centrée à l'origine de l'espace gaussien et de rayon  $\beta_{HL}$  (indice de Hasofer-Lind défini par la formule (1.2.2)). Nous noterons cette hyper-sphère  $S_{\beta_{HL}}$ . Nous pouvons constater qu'il n'existe pas d'autres points que le point de conception appartenant à cette hyper-sphère et au domaine de défaillance, une illustration à été donnée dans le chapitre précédent (1.2.3).

La probabilité de défaillance peut s'écrire :

$$\begin{aligned} P_f &= P\left(\{H(u) \leq 0\} \cap \{u \in S_{\beta_{HL}}^c\}\right) \\ &= P(H(u) \leq 0 | u \in S_{\beta_{HL}}^c) \cdot P(u \in S_{\beta_{HL}}^c), \end{aligned} \quad (2.2.1)$$

où  $P(u \in S_{\beta_{HL}}^c) = 1 - \chi_d^2(\beta_{HL}^2)$  avec  $\chi_d^2(\cdot)$  loi du Chi-2 à  $d$  degrés de liberté. Si cette quantité est trop importante (dimension  $d$  grande ou  $\beta_{HL}$  faible), la méthode devient peu efficace.

Dans [SM 93], il est proposé de calculer  $P(H(u) \leq 0 | u \in S_{\beta_{HL}}^c)$  par tirage d'importance utilisant comme densité d'importance une densité multinormale, dont les composantes sont indépendantes, centrée au point de conception et réduite. Une illustration du tirage d'importance conditionné est donnée figure (2.2.1).

Cette technique est efficace à partir du moment où le point de conception est unique et déterminé précisément. Si la sphère est trop petite, on perd seulement de l'efficacité, mais si la sphère est trop grande, la probabilité de défaillance est sous-estimée (l'estimateur de la probabilité de défaillance peut devenir biaisé). Enfin, l'efficacité de cette méthode, gagnée en excluant une sphère, décroît quand la dimension  $d$  du problème augmente.

### 2.2.1.3 Tirage d'importance asymptotique

Cette approche [MBD 93] est basée sur les résultats de Breitung dans [Bre 84]. Elle permet la construction d'une densité d'importance qui tend vers la densité optimale  $f^*(x) = f_X(x|x \in D)$  quand  $P_f \rightarrow 0$ , où  $D$  est le domaine de défaillance. Le point de départ de cette méthode est d'effectuer une transformation de l'espace des variables d'entrées telle que :

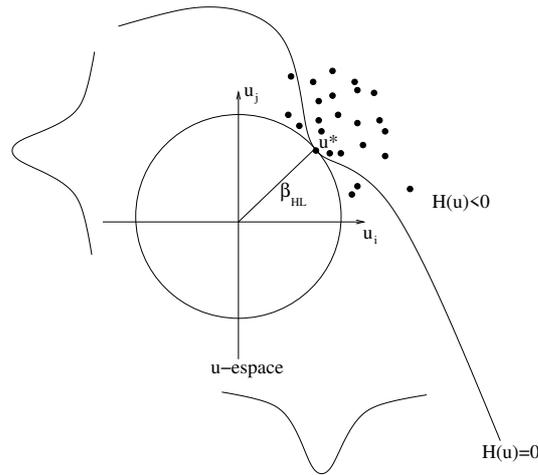


FIG. 2.2.1 – Tirage d'importance conditionné

1. l'origine du nouveau repère soit le point de conception sur la surface d'état limite
2. le premier axe suive la direction de la descente la plus raide de la surface d'état limite ;
3. les autres directions soient déterminées à partir des courbures principales de la fonction d'état limite au point de conception.

La première variable de la densité d'importance est indépendante des autres et est distribuée selon une loi exponentielle de moyenne  $1/|\nabla L|_{x=x^*}$  où  $L = \ln f_X(x)$ , les  $d - 1$  autres sont distribuées selon une loi multinormale centrée et de matrice de covariance  $C^{-1}$  :

$$C = \left\{ \frac{\partial^2 L}{\partial x_i \partial x_j} - \frac{|\nabla L|}{|\nabla G|} \frac{\partial^2 G}{\partial x_i \partial x_j} \right\}_{i,j=2,\dots,d}.$$

Cette méthode demande malheureusement le calcul de la matrice Hessienne de  $G$  au point de conception, et est ainsi difficilement applicable en grande dimension.

#### 2.2.1.4 Mise à jour de la probabilité approchée

Dans [HR 88, FR 88], les auteurs considèrent l'utilisation du tirage d'importance dans la mise à jour de l'estimation de la probabilité de défaillance déterminée par la méthode SORM. Dans [ER 93], les mêmes corrections sont faites mais avec un calcul FORM. La mise à jour est faite en introduisant un facteur de correction  $C$ . La probabilité de défaillance s'exprime par :

$$P_f = P_{approx} \frac{P(H(X) \leq 0)}{P_{approx}} = P_{approx} \cdot C, \quad (2.2.2)$$

où  $P_{approx}$  est la probabilité de défaillance calculée par FORM ou SORM et  $\{H(X) \leq 0\}$  le domaine de défaillance dans l'espace gaussien.

Nous nous plaçons dans l'espace gaussien tel que le point de conception  $u^*$  ait pour coordonnées  $(0, 0, \dots, \beta_{HL})$  (comme pour la simulation axis-orthogonale), cf § 1.3.9.4. Le vecteur  $U$  se décompose en  $U = (\tilde{U}, U_d).y$

Pour le résultat FORM,  $C$  s'exprime par :

$$C \approx \frac{1}{n} \sum_{i=1}^n \frac{\Phi(-\hat{u}_d^i)}{\Phi(-\beta_{HL})} \quad (2.2.3)$$

$\beta_{HL}$  est l'indice de Hasofer-Lind,  $\hat{u}_d$  la racine de  $H_{axis}(\tilde{u}, u_d) = 0$  et  $\tilde{u} = (u_1, \dots, u_{d-1})$  un vecteur dont les composantes sont indépendantes, normales, centrées et réduites.

Pour le résultat SORM, nous nous servons des courbures de la fonction d'état limite au point de conception,  $C$  s'exprime par :

$$C \approx \frac{1}{n} \sum_{i=1}^n \frac{\Phi(-\hat{u}_d^i)}{\Phi(-\beta_{HL})} \exp \left[ \frac{1}{2} \frac{\Phi(-\hat{u}_d^i)}{\Phi(-\beta_{HL})} \sum_{j=1}^{d-1} k_i \tilde{u}_i^j \right] \quad (2.2.4)$$

$\beta_{HL}$  l'indice de Hasofer-Lind,  $k_i$  les courbures principales au point de conception,  $u_d$  est la racine de  $H_{axis}(\tilde{u}, u_d) = 0$  et  $\tilde{u} = (u_1, \dots, u_{d-1})$  un vecteur dont les composantes sont indépendantes, normales, centrées et de variance  $\text{Var}[\tilde{u}_i] = \left(1 + \frac{\Phi(-\hat{u}_d^i)}{\Phi(-\beta_{HL})} k_i\right)^{-1}$ . Cela implique des conditions sur les courbures  $k_i$  si l'on veut que  $\text{Var}[\tilde{u}_i]$  soit finie.

Que ce soit pour FORM ou SORM, chaque simulation demande le calcul de  $u_d$ , la racine de  $H_{axis}(\tilde{u}, u_d) = 0$ , c'est-à-dire un surcoût en nombre d'appel à la fonction d'état limite. De plus, la calcul de la hessienne dans SORM peut être trop coûteux.

### 2.2.1.5 Approche adaptative de Bucher

Le principe de cette méthode [Buc 88] est de s'approcher, de manière adaptative, des moments de la densité optimale conditionnellement au domaine de défaillance donnée par la formule (2.1.7).

La démarche est la suivante :

1. Étape initiale : Génération d'un échantillon  $x^{(0)}$  suivant la densité initiale. A partir de cet échantillon, on effectue le calcul des moments conditionnellement au domaine de défaillance :

$$\mu_0 = \mathbb{E}[x^{(0)} | x^{(0)} \in D] \text{ et } \sigma_0^2 = \text{Var}(x^{(0)} | x^{(0)} \in D).$$

2. Étape  $i$  : génération d'un échantillon suivant une loi multinormale dont les paramètres ont été déterminés précédemment. Puis on calcule à nouveau les moments du nouvel échantillon conditionnellement au domaine de défaillance.

L'étape initiale peut ne pas permettre de générer des points dans le domaine de défaillance, il est préférable d'utiliser la densité initiale centrée au point de conception (ce qui implique la recherche de ce dernier), cela favorisera les tirages dans le domaine de défaillance. De plus, si la variance estimée est trop petite, il peut être utile de l'augmenter.

### 2.2.1.6 Approche adaptative de Karamchandani, Bjerager et Cornell

L'idée de cette méthode adaptative [KBC 88] est de construire une densité d'importance à chaque étape en utilisant une densité multimodale et un ensemble de points représentatifs.

La démarche est la suivante :

1. Étape initiale : génération d'un échantillon de points suivant la densité initiale mais centrée en un point  $x^{(0)}$  appartenant au domaine de défaillance (ce point est choisi arbitrairement par l'utilisateur). Puis on identifie les points représentatifs  $(\hat{x}^{(1)}, \dots, \hat{x}^{(k)})$  (l'identification de cet ensemble de points est expliqué plus loin).
2. Étape  $i$  : génération des nouveaux tirages par une densité multimodale construite à partir des points représentatifs déterminés précédemment :

$$\tilde{f}^{(i)}(x) = \sum_{j=1}^k \frac{f_X(\hat{x}^{(j)})}{\sum_{l=1}^k f_X(\hat{x}^{(l)})} f_X^{(j)}(x),$$

où  $f_X^{(j)}$  est la densité initiale centrée en  $\hat{x}^{(j)}$  :  $f_X^{(j)}(x) = f_X(x - \hat{x}^{(j)})$  et  $f_X$  la densité initiale. Puis on recherche les nouveaux points représentatifs.

*Identification des points représentatifs :*

Soit  $S_i$  un ensemble de points du domaine de défaillance, on identifie le point de plus grande probabilité  $\hat{x}^{(1)}$ , on élimine ensuite les points de  $S_i$  appartenant à la sphère centrée en  $\hat{x}^{(1)}$  et de rayon  $d_0$  fixé, parmi l'ensemble des points restants de  $S_i$  on détermine le point de plus grande probabilité  $\hat{x}^{(2)}$ , et ainsi de suite jusqu'à ce qu'on ait éliminé tous les points de  $S_i$ . On se retrouve alors avec  $k$  points représentatifs  $(\hat{x}^{(1)}, \dots, \hat{x}^{(k)})$ .

La probabilité de défaillance est calculée par :

$$\hat{P}_f = \frac{1}{n} \sum_{i=1}^n \frac{f_X(X_i)}{\tilde{f}^{(i)}(X_i)} \mathbf{1}_{\{G(X_i) \leq 0\}}(X_i), \quad (2.2.5)$$

$$\text{avec } \tilde{f}^{(i)}(x) = \sum_{j=1}^k \frac{f_X(\hat{x}^{(j)})}{\sum_{l=1}^k f_X(\hat{x}^{(l)})} f_X^{(j)}(x), \quad (2.2.6)$$

où  $f_X^{(j)}$  est la densité initiale centrée en  $\hat{x}^{(j)}$  et  $(\hat{x}^{(1)}, \dots, \hat{x}^{(k)})$  les points représentatifs du tirage  $i - 1$ .

En augmentant le nombre de points dans le domaine de défaillance, on tend vers un ensemble dont les points représentatifs sont équidistants et qui indiquent les régions importantes du domaine de défaillance. Dès lors, on peut réduire  $d_0$  et diminuer la variance de  $f_X^{(j)}$ .

Cette méthode peut s'avérer coûteuse en nombre d'appels à la fonction de défaillance, à chaque étape, la détermination de l'ensemble  $S_i$  demande autant d'appels à  $G$  que de tirages effectués.

### 2.2.1.7 Technique adaptative de Wu

La technique itérative de Wu [Wu 92] suggère de perturber la surface d'approximation quadratique de SORM à chaque itération. On se place dans l'espace gaussien standard, puis on effectue une rotation de celui-ci. Au départ, la surface quadratique est une surface parabolique de la forme suivante :

$$H_s = \beta_{HL} - v_n + \frac{1}{2} \sum_{i=1}^{n-1} k_i v_i^2, \quad (2.2.7)$$

où  $v$  est un vecteur de loi multinormale centrée réduites dont les composantes sont indépendantes dans le nouvel espace (espace gaussien + rotation telle que le point de conception  $v^*$  ait pour coordonnées  $(0, 0, \dots, \beta_{HL})$ ), et  $k_i$  sont les courbures principales de la fonction d'état limite au point de conception. La démarche est ensuite la suivante :

1. Génération d'un échantillon  $U$  dans le domaine  $S$  définie par  $\{H_s \leq 0\}$ . On utilise un échantillon de petite taille pour calculer  $\hat{P}$  défini par :

$$\hat{P} = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{1}_{\{H(U_i) \leq 0\}},$$

où  $H$  représente la surface d'état limite. La taille de l'échantillon  $n_1$  dans  $S$  peut-être calculée par un intervalle de confiance de seuil  $1 - \alpha$  (en supposant que  $\hat{P}$  est gaussien) et une erreur  $\gamma$  :

$$\gamma = \phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \sqrt{\frac{1 - \hat{P}}{\hat{P} n_1}}.$$

2. Perturbation de la fonction parabolique en changeant les courbures au point de conception de manière à ce que  $P_{S'}$  ( $P_{S'}$  est la probabilité de la région  $S'$ ) soit supérieur à  $P_S$  ( $S' = S + \Delta S$ ). La nouvelle surface parabolique est :

$$H_{s'} = \beta_{HL} - v_n + \frac{1}{2} \sum_{i=1}^{n-1} k'_i v_i^2.$$

Le nombre de tirage dans  $\Delta n$  est donné par :

$$\Delta n = \frac{P_{S'} - P_S}{P_S} \cdot n_1$$

où  $P_S$  est la probabilité du domaine  $S$  et où  $P_{S'}$  la probabilité du domaine  $S'$ .

3. Répétition de l'étape 2 jusqu'à ce qu'il n'y ait plus de point dans le domaine de défaillance et jusqu'à ce que  $\gamma$  soit assez petit.
4. La probabilité de défaillance devient :

$$P_f = P_{S'} P \left( H(U) \leq 0 \mid U = u \in S' \right).$$

Une approche similaire est aussi présentée dans l'article [Mel 99] pour la fiabilité des systèmes.

### 2.2.1.8 Méthodes utilisant des noyaux

Nous allons présenter ici trois méthodes de construction de densité d'importance à partir de la méthodes des noyaux, l'une directe [AAT 89, AAT 92] et les autres adaptatives [WAL 94, AB 99].

#### Méthode directe

On commence par générer un échantillon de taille  $k$  suivant la densité initiale, ensuite on estime la densité d'importance par :

$$\tilde{f}(x) = \frac{1}{j} \sum_{i=1}^j \frac{1}{w^d} K_d \left( \frac{x - X_i}{w} \right), \quad (2.2.8)$$

avec :

- $j$  le nombre de tirages effectués dans le domaine de défaillance,
- $K_d$  le noyau gaussien en dimension  $d$ ,
- $w$  est la taille de la fenêtre de lissage. On peut adapter la taille de la fenêtre en utilisant  $\lambda_i w$  qui permet d'adapter la taille de la fenêtre à la densité des régions (plus grande pour une région à faible densité) :  $\lambda_i = \left( \frac{1}{j} \sum_{i=1}^j \frac{1}{w^d} K_d \left( \frac{x - X_i}{w} \right) \right)^{-\alpha}$ ,  $\alpha \in [0, 1]$ .

Le problème de cette méthode réside dans la génération de l'échantillon initial, il nous faut des points dans le domaine de défaillance, or si la probabilité cherchée est faible, il nous faudra un nombre beaucoup trop important de tirages, ce qui rend impossible l'application de cette méthode. En plus, pour estimer efficacement une densité par la méthode du noyau, il faut disposer d'un nombre conséquent de points.

#### Méthodes adaptatives

Dans les méthodes adaptatives, c'est surtout la manière de générer l'échantillon de départ, et donc les  $j$  points du domaine de défaillance, qui diffère de la méthode directe. Idéalement, on cherche à générer un échantillon suivant  $f^*$  donnée par (2.1.7).

Dans [WAL 94], la génération de l'échantillon initial se fait avec une loi multinormale centrée au point de conception. Comme dans l'approche de Bucher (cf. § 2.2.1.5), les moments de l'échantillon conditionnellement au domaine de défaillance sont calculés et servent à la construction du noyau gaussien.

Dans [AB 99], il est proposé d'utiliser les méthodes de Monte Carlo par chaînes de Markov, et notamment l'algorithme de Metropolis-Hasting, (cf. § 3.2.1) qui permettent d'obtenir une chaîne de Markov dont les réalisations sont asymptotiquement distribuées selon  $f^*$ .

### 2.2.1.9 Tirage d'importance filtré avec SVM

Récemment, une nouvelle méthode utilisant les machines à vecteurs supports (SVM) a été développée [Hur 07]. Pour une brève introduction aux SVM dans un contexte de classification, on peut se référer à [Bur 98], cette méthode permet de construire une surface discriminante à partir d'un échantillon d'apprentissage. Ainsi, à partir d'un échantillon dont on connaît l'appartenance ou non au domaine de défaillance, on construit une surface qui devrait approcher la surface d'état limite. La méthode du tirage d'importance est ensuite utilisée pour le calcul de la probabilité de défaillance.

La démarche est la suivante :

1. Génération d'un échantillon de taille  $n$  suivant la densité d'importance  $\tilde{f}$ , qui peut être par exemple une densité multinormale centrée au point de conception.
2. – i) Initialisation de la probabilité de défaillance à 1 :  $\hat{P}_f^0 = 1$ .
  - ii) Sélection d'un petit ensemble de points, dont on détermine l'appartenance ou non au domaine de défaillance.
  - iii) Approximation de la surface d'état limite par les SVM, on note cette approximation  $\{\hat{G}(X) = 0\}$ .
  - iv) Calcul de la probabilité de défaillance sur l'échantillon restant (échantillon de départ moins ensemble d'apprentissage) en utilisant  $\{\hat{G}(X) = 0\}$  comme surface d'état limite. On compare alors la probabilité de défaillance obtenue  $\hat{P}_f^k$  avec la précédente  $\hat{P}_f^{k-1}$ , si  $|\hat{P}_f^k - \hat{P}_f^{k-1}| < \varepsilon$  ( $\varepsilon$  arbitraire), on arrête la construction de  $\{\hat{G}(X) = 0\}$ , sinon, on augmente la taille de notre petit ensemble de points de l'étape ii) en sélectionnant des points dans la marge (entre les surfaces  $\{\hat{G}(X) = -\delta\}$  et  $\{\hat{G}(X) = \delta\}$ ,  $\delta$  fixé arbitrairement) et on retourne à l'étape iii).
3. L'estimateur final de la probabilité de défaillance est alors :

$$\hat{P}_f = \frac{1}{n} \left\{ \sum_{i=1}^j \mathbf{1}_{\{G(X_i) \leq 0\}} \frac{f_X(X_i)}{\tilde{f}(X_i)} + \sum_{i=1}^{n-j} \mathbf{1}_{\{\hat{G}(X_i) \leq 0\}} \frac{f_X(X_i)}{\tilde{f}(X_i)} \right\},$$

où  $j$  est le nombre de tirages utilisés pour construire  $\{\hat{G}(X) = 0\}$ .

En conclusion, le tirage d'importance aide la construction de  $\{\hat{G}(X) = 0\}$  en concentrant les tirages dans les zones d'intérêts (on a plus de tirages dans le domaine de défaillance), et l'approximation de la surface d'état limite par les SVM allège le nombre d'évaluations de la fonction  $G$ . Les méthodes d'apprentissage statistique de discrimination peuvent servir à la construction de la fonction d'état limite, on peut se référer à [Hur 04].

### 2.2.1.10 Tirage d'importance pour la fiabilité des systèmes

Pour une introduction à la fiabilité des systèmes, on peut se reporter au paragraphe 1.2.2.5. Nous allons décrire brièvement l'approche de R. E Melchers [Mel 99].

Pour un système série constitué de  $m$  composants ( $m$  fonctions d'état limite) la densité d'importance est une densité multimodale :

$$\tilde{f} = \sum_{i=1}^m \alpha_i \tilde{f}^{(i)}, \quad \sum_{i=1}^m \alpha_i = 1,$$

où  $\tilde{f}^{(i)}$  est la densité d'importance correspondant à l'état limite  $i$ , déterminée par une des procédures présentées précédemment (densité initiale centrée au point de conception par exemple).

Pour un système parallèle constitué de  $m$  composants, les régions importantes sont proches des intersections des fonctions d'état limite, une fois les  $k$  points d'intersections déterminés (deux fonctions d'état limite peuvent avoir plusieurs points d'intersections), on peut construire une densité d'importance multimodale dont les  $k$  composante seront centrées en ces points.

### Tirage d'importance multimodal tronqué

C'est une approche pour la fiabilité des systèmes [Fu 94] dans le cas où les variables  $X$  de départ sont distribuées selon une loi gaussienne  $f$  de moyenne  $\bar{x}$  et de matrice de covariance  $C$ . On considère qu'il y a  $m$  modes de défaillances, et on introduit alors la fonction d'état limite  $G_i$  du mode  $i$ . Cette méthode présente la construction d'une densité d'importance pour des problèmes dont les surfaces d'état limite sont linéaires ou linéarisées et ainsi de la forme :

$$a_i^t(x - x_i^*) = 0, \quad a_i = \nabla G_i(x)|_{x=x_i^*},$$

où  $x_i^*$ , est le point de conception pour le mode  $m$  de défaillance. La densité d'importance suggérée est :

$$\tilde{f}(x) = \sum_{i=1}^m w_i f_i(x), \quad (2.2.9)$$

où  $f_i(x)$  est la densité multinormale de moyenne  $x_i^*$ , tronquée par l'hyperplan  $\{a_i^t(x - x_i^*) = 0\}$  et de covariance  $C$ . Les poids  $w_i$  sont déterminés en résolvant :

$$\frac{f(x_1^*)}{f(x_i^*)} = \frac{\tilde{f}(x_1^*)}{\tilde{f}(x_i^*)}, \quad i = 2, \dots, m \text{ et } \sum_{i=1}^m w_i = 1, \quad w_i \geq 0, \quad i = 1, \dots, m. \quad (2.2.10)$$

La probabilité de défaillance est estimée par :

$$\hat{P}_f = \sum_{i=1}^M \frac{w_i}{N_i} \left\{ \sum_{j=1}^{N_i} \mathbf{1}_{\{G(X_j^i) \leq 0\}} \frac{f(X_j^i)}{\tilde{f}(X_j^i)} \right\}, \quad X_i^i \sim f_i, \quad (2.2.11)$$

avec  $N_i \approx w_i N$  et  $\sum_{i=1}^m N_i = N$ . La variance de  $\hat{P}_f$  est majorée par  $\frac{1}{N} \left\{ P_f \frac{f(x^*)}{\tilde{f}(x^*)} - P_f^2 \right\}$ , où  $x^*$  est l'un des  $m$  points de conception.

### 2.2.1.11 Cas fiabilité des systèmes et calcul FORM

Pour calculer la probabilité de défaillance d'un système (série ou parallèle) donné par les formules (1.2.14) et (1.2.15), il nous faut évaluer une fonction de répartition multinormale de dimension  $d$  égale au nombre de modes de défaillance. Or pour une dimension supérieure à 7, c'est infaisable. Des solutions ont été proposées par [GR 88] ou par [Joe 95]. Nous allons en présenter deux qui utilisent le tirage d'importance.

Rappelons (cf. § 1.2.2.5) que l'on cherche à calculer  $\hat{P}_f = 1 - \Phi_d(\beta, R)$  dans le cas d'un système série et  $\hat{P}_f = \Phi_d(-\beta, R)$  dans le cas d'un système parallèle, avec  $\Phi_d(\beta, R)$  fonction de répartition d'une loi multinormale centrée de dimension  $d$  de matrice de corrélation  $R$  ( $r_{ij} = a_i^t a_j, r_{ii} = 1$ ) et évaluée en  $\beta = (\beta_{HL}^1, \dots, \beta_{HL}^d)^t$ .

#### Tirage d'importance séquentiel conditionné

C'est une méthode [ADKOS 98] utilisée pour le calcul de :

$$I = \int_{\prod_{i=1}^d [a_i, b_i]} \varphi(x_1, \dots, x_d) dx_1 \cdots dx_d, \quad (2.2.12)$$

$\varphi$  densité d-multinormale de moyenne  $m = (m_1, \dots, m_d)$  et de matrice de covariance  $R$  d'éléments  $r_{ij}, i, j = 1, \dots, d$ .

Nous présenterons cette méthode dans un contexte général (nous n'avons pas forcément  $r_{ii} = 1$ ).

La démarche est donnée par :

1. Étape 1 : Génération d'une valeur  $x_1$  suivant la densité :

$$\frac{\varphi_1(x_1)}{\Phi_1([a_1, b_1])} \mathbf{1}_{\{[a_1, b_1]\}}(x_1),$$

où  $\Phi_1([a_1, b_1]) = \Phi_1(a_1) - \Phi_1(b_1)$ ,  $\varphi_1$  et  $\Phi_1$  respectivement la densité et la fonction de répartition de la loi normale de moyenne  $m_1$  et de variance  $r_{11}$ .

2. Étape d : Après avoir obtenu  $x_1, \dots, x_{d-1}$ , génération d'une valeur suivant la densité :

$$\frac{\varphi_d(x_d|x_1, \dots, x_{d-1})}{\Phi_d([a_d, b_d]|x_1, \dots, x_{d-1})} \mathbf{1}_{\{[a_d, b_d]\}}(x_d),$$

où  $\varphi_d$  et  $\Phi_d$  respectivement la densité et la fonction de répartition de la densité normale de moyenne  $m_d(x_1, \dots, x_{d-1})$  et de variance  $v_d$  :

$$m_d(x_1, \dots, x_{d-1}) = m_d - \sum_{i=1}^{d-1} c_{di} \frac{x_i - m_i}{c_{dd}}, \quad v_d = r_{dd},$$

où les  $c_{ij}$  sont les éléments de la matrice  $C = R^{-1}$ .

3. La probabilité de défaillance est alors approchée par  $\hat{P}_f = \frac{1}{n} \sum_{i=1}^n Y_i$ , où  $Y_i$  est une réalisation de la variable aléatoire  $Y = \prod_{k=1}^d \Phi_k([a_k, b_k]|x_1, \dots, x_{k-1})$ , obtenue à partir d'un échantillon  $(x_1, \dots, x_d)$  construit durant les précédentes étapes.

En effet, nous avons :

$$\begin{aligned} \mathbb{E}[Y] &= \int_{\prod_{i=1}^d [a_i, b_i]} \prod_{k=1}^d \Phi_k([a_k, b_k]|x_1, \dots, x_{k-1}) \\ &\quad \times \prod_{k=1}^d \frac{f_k(x_k|x_1, \dots, x_{k-1})}{\Phi_k([a_k, b_k]|x_1, \dots, x_{k-1})} \mathbf{1}_{\{[a_k, b_k]\}}(x_k) dx_1 \cdots dx_d \\ &= \int_{\prod_{i=1}^d [a_i, b_i]} \varphi_1(x_1) \varphi_2(x_2|x_1) \cdots \varphi_d(x_d|x_1, \dots, x_{d-1}) dx_1 \cdots dx_d \\ &= I \end{aligned}$$

Il est important de noter que l'efficacité de cette méthode décroît quand la dimension  $d$  augmente et les corrélations entre les variables augmentent.

En posant  $m_i = 0$ ,  $a_i = -\infty$  et  $b_i = \beta_{HL}^i$  pour un système série et  $b_i = -\beta_{HL}^i$  pour un système parallèle, on se retrouve bien dans la situation voulue. De plus, dans le cas très particulier où le domaine de défaillance est définie par  $\prod_{i=1}^m [a_i, b_i]$  et les variables d'entrée sont distribuées selon une loi multinormale, on peut appliquer directement cette méthode pour le calcul de la probabilité de défaillance.

Dans [PS 02], l'approche du tirage d'importance séquentiel conditionné pour le calcul de  $\Phi_m(\beta, R)$  est comparée à une simple méthode d'approximation (méthode PCM : *product of conditional marginals*) où  $\Phi_m(\beta, R) \approx \prod_{k=1}^m \Phi(\beta_k|_{k-1}) = P\left((X_k \leq \beta_k) \mid \bigcap_{i=1}^{k-1} (X_i \leq \beta_i)\right)$ . Il s'avère qu'en grande dimension et en présence de fortes corrélations, ces deux méthodes ont une efficacité comparable mais la méthode PCM est plus simple à mettre en oeuvre.

### Autre approche

Dans [MK 03], il est proposé de calculer directement  $1 - \Phi_m(\beta, R) = \Phi_m^C(\beta, R)$  (système série) par une technique de tirage d'importance. Rappelons que dans l'espace gaussien standard de dimension  $m$ , la fonction de défaillance linéarisée du mode  $k$  est donnée par  $H_k^*(u) = a_k^t u + \beta_k$  et la probabilité de défaillance de chaque mode est approchée par  $\Phi(-\beta_i)$ . Un estimateur de  $\Phi_m^C(\beta, R)$  est donné par :

$$\hat{\Phi}_m^C(\beta, R) = \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{1}_{\{\cup_{k=1}^m H_k^*(V_i) \leq 0\}} \prod_{k=1}^m \varphi(V_{ik})}{h_V(V_k)}, \text{ où} \quad (2.2.13)$$

- $V_i = (V_{i1}, \dots, V_{im})^t \sim h_V$ ,
- $h_V(u) = \sum_{k=1}^m w_k h_V^k(u)$ ,
- $h_V^k(u) = \frac{\mathbf{1}_{\{H_k^*(u) \leq 0\}} \prod_{k=1}^m \varphi(u_j)}{\Phi(-\beta_i)}$
- $w_i = \Phi(-\beta_i) / (\sum_{k=1}^m \Phi(-\beta_i))$ .

### 2.2.1.12 Comparaison de l'efficacité des méthodes

Certaines de ces méthodes peuvent aussi être adaptées dans des problèmes de fiabilité des structures avec variation du temps, on peut voir par exemple [Rac 01].

Certains auteurs se sont attachés à comparer l'efficacité (nombre d'appels à la fonction d'état limite et précision) de certaines des méthodes présentées dans cette section [ER 93] et plus récemment [ETL 05]. Cette comparaison peut se faire soit en comparant le nombre d'appels à la fonction d'état limite pour une précision fixée (coefficient de variation égal à 5% par exemple), soit en comparant la variance des différents estimateurs pour un nombre d'appel à la fonction d'état limite fixe. Des tests sont effectués sur plusieurs exemples de fonctions d'état limite (un point de conception, plusieurs points de conception, système série, système parallèle,...); ces articles présentent quelques recommandations sur le choix des méthodes basées sur certains critères (ressource informatiques disponibles, précision demandée, caractéristiques de la fonction d'état limite, etc...)

S.K. Au et J.L. Beck s'intéressent dans [AB 03] à l'applicabilité (différent de l'efficacité) du tirage d'importance en grande dimension :

*Le tirage d'importance est applicable en grande dimension si, pour un problème de fiabilité des structures donné  $\mathcal{P}_{d_k}$  de dimension  $d_k$ , et une densité  $\tilde{f}_{d_k}$  appartenant à une famille de densités donnée  $\mathcal{F}_{d_k}$ , nous avons*

$$\text{Var}(\hat{P}_f) < +\infty \text{ quand } k \rightarrow +\infty,$$

avec  $\{d_k \in \mathbf{Z}^+, k = 1, 2, \dots\}$ ,  $d_k \rightarrow +\infty$  quand  $k \rightarrow +\infty$

Par exemple, pour un problème dont les paramètres sont distribués selon une gaussienne centrée réduite de dimension  $d$ , et une densité d'importance appartenant à la famille des densités multivariées de dimension  $d$  centrées au point de conception  $x^*$  et de matrice de covariance  $C$ , dont les valeurs propres sont supérieures à  $1/2$ , le tirage d'importance est applicable en grande dimension si et seulement si :

1. il y a au plus un nombre fini de valeurs propres de  $C$  égales à 1 quand  $d \rightarrow +\infty$  ;
2.  $\|x^*\| < +\infty$  quand  $d \rightarrow +\infty$ .

## 2.2.2 Construction de quelques densités d'importance dans un cadre général

Nous commencerons par faire quelques considérations sur le rapport de vraisemblance  $w$  car son comportement est à la base d'un choix judicieux de la densité d'importance. Ensuite nous verrons quelques stratégies de construction de la densité d'importance. Rappelons que  $w(x) = \frac{f(x)}{\tilde{f}(x)}$ .

### 2.2.2.1 Le rapport de vraisemblance

Notons tout d'abord que l'espérance mathématique de  $w$  est 1 :  $\mathbb{E}_{\tilde{f}}[w(X)] = 1$ .

Nous commencerons par introduire quelques points importants sur la densité d'importance et sur le rapport de vraisemblance :

1. On cherche une densité d'importance à partir de laquelle il est facile de générer des variables aléatoires.
2. Le rapport de vraisemblance doit être facile à calculer.
3. On cherche  $\tilde{f}$  telle que la variance de  $\hat{I}$  soit finie. Or, la variance de  $\hat{I}$  (équation (2.1.1)) est finie si :

$$\mathbb{E}_{\tilde{f}} [g^2(X)w^2(X)] = \mathbb{E}_f [g^2(X)w(X)] < +\infty. \quad (2.2.14)$$

On peut remarquer qu'une densité d'importance  $\tilde{f}$  dont les queues sont à décroissance plus rapide que celles de la densité initiale  $f$  n'est pas appropriée,  $\text{Var}(\hat{I})$  sera infinie pour de nombreuses fonctions  $g$ .

4. On cherche une densité d'importance qui diminue la variance de l'estimateur par rapport à une estimation par la méthode de Monte Carlo directe, c'est-à-dire que l'on cherche  $\tilde{f}$  telle que :

$$\text{Var}(\hat{I}^{MC}) - \text{Var}(\hat{I}^{TI}) = \frac{1}{n} \mathbb{E}_f [g^2(X) (1 - w(X))] \geq 0 \quad (2.2.15)$$

Cela nous indique que l'on cherche une densité d'importance de telle sorte que le rapport de vraisemblance  $w$  soit grand quand  $g$  est proche de 0 ( $w$  petit quand  $g^2$  est grand).

5. La forme de la densité optimale (2.1.5), suggère de chercher une densité d'importance telle que  $|g|w$  soit pratiquement une constante et de variance finie. De plus, il est important de noter qu'un choix de densité d'importance  $\tilde{f}$  telle  $\mathbb{E}_f[w] = +\infty$ , n'est pas recommandé [RC 99] : le comportement de l'estimateur devient alors instable.

Pour satisfaire le point 3), J. Geweke [Gew 89] présente deux conditions suffisantes :

- $w(x) < M, \forall x \in D$  et  $\text{Var}_f(g(X)) < +\infty$ .
- $f(x) < F$  et  $\tilde{f} > \varepsilon \forall x \in D$  ensemble compact.

Comme on ne choisit pas le domaine d'intégration  $D$ , la deuxième condition ne peut être satisfaite que dans des cas vraiment particuliers.

T. C. Hesterberg dans [Hes 88] mentionne lui aussi un choix de densité d'importance  $\tilde{f}$  telle que le rapport de vraisemblance soit borné  $w(x) < M$ . La variance de  $w$  est alors majoré :  $\text{Var}_{\tilde{f}}(w(X)) \leq M - 1$ . Si le rapport de vraisemblance n'est pas borné, on risque de donner une trop grande importance à une petite fraction de l'échantillon. Générer un échantillon suivant  $\tilde{f}$  telle que  $w(x) < M, M \geq 1$  est facilité par la possibilité d'employer la méthode de rejet (on peut même voir la méthode de rejet comme un tirage d'importance avec un choix particulier de densité d'importance [Che 05]).

On obtient aussi une comparaison entre les variances des estimateurs  $\hat{I}$  et  $\hat{I}'$  (équation (2.1.8) avec la variance de l'estimateur d'un Monte Carlo direct :

$$\text{Var}(\hat{I}) \leq M\text{Var}(\hat{I}^{MC}) - (1 - M)I^2, \text{ et} \quad (2.2.16)$$

$$\text{Var}(\hat{I}') \leq M\text{Var}(\hat{I}^{MC}). \quad (2.2.17)$$

Il en découle alors une condition suffisante de réduction de variance d'un facteur  $\rho$  ( $\rho \leq 1$ ) par rapport à une méthode de Monte Carlo directe :

$$\text{Si } w(x) \leq \rho \text{ quand } g(x) \neq 0,$$

$$\text{alors } \text{Var}(\hat{I}) \leq \rho \text{Var}(\hat{I}^{MC}), \quad (2.2.18)$$

$$\text{et } \text{Var}(\hat{I}') \leq \rho \text{Var}(\hat{I}^{MC}). \quad (2.2.19)$$

Cette condition est très restrictive pour le choix de  $\tilde{f}$  et pas vraiment utilisable en pratique.

### 2.2.2.2 Méthodes non paramétriques

Les méthodes non paramétriques utilisent les méthodes de noyaux (cf. §2.2.1.8). Dans [GR 96, Zha 96] des méthodes adaptatives non paramétriques à  $k$  étapes sont proposées. La démarche est la suivante :

1. Étape 0 : génération d'un échantillon  $X_1^{(0)}, \dots, X_{n_0}^{(0)}$  suivant une densité d'importance initiale arbitraire  $\tilde{f}_0$  et on calcule la nouvelle densité d'importance par :

$$\tilde{f}_1(x) = \frac{1}{\sum_{i=1}^{n_0} w_0(X_i^{(0)})} \sum_{i=1}^{n_0} \frac{w_0(X_i^{(0)})}{(h^{(0)})^d} K_d \left( \frac{x - X_i^{(0)}}{h^{(0)}} \right)$$

2. Étape k : génération d'un échantillon  $X_1^{(k-1)}, \dots, X_{n_{k-1}}^{(k-1)}$  la densité d'importance  $\tilde{f}_{k-1}$  et on calcule la nouvelle densité d'importance par :

$$\tilde{f}_k(x) = \frac{1}{\sum_{i=1}^{n_{k-1}} w_{k-1}(X_i^{(k-1)})} \sum_{i=1}^{n_{k-1}} \frac{w_{k-1}(X_i^{(k-1)})}{(h^{(k-1)})^d} K_d \left( \frac{x - X_i^{(k-1)}}{h^{(k-1)}} \right),$$

où  $K_d(u_1, \dots, u_d) = K(u_1) \cdots K(u_d)$  et  $K$  un noyau unidimensionnel.

Il existe un  $h^{(k-1)}$  optimal qui minimise l'erreur quadratique moyenne asymptotique de  $\hat{I}_k$  (estimateur de  $I$  par tirage d'importance avec la densité d'importance  $\tilde{f}_k$ ), il peut être estimé [Zha 96]. Il apparaît aussi qu'à nombre de simulations fixé, le tirage d'importance non paramétrique non adaptatif ( $k = 0$ ) est plus efficace que tirage d'importance non paramétrique adaptatif ( $k \geq 0$ ). Cette méthode peut être une alternative intéressante au tirage d'importance quand celui-ci échoue. Cependant, le principale problème de cette méthode reste qu'il faut un échantillon de grande taille pour estimer une densité par la méthode du noyau.

### 2.2.2.3 Approche par mélange

L'idée est d'utiliser une densité d'importance  $\tilde{f}$  comme une combinaison linéaires d'autres densités :

$$\tilde{f}(x) = \sum_{i=1}^k \lambda_i \tilde{f}_i(x), \quad \sum_{i=1}^k \lambda_i = 1 \text{ et } \lambda_i \geq 0. \quad (2.2.20)$$

#### Approche de Oh et Berger

Cette approche [BO 93] consiste à choisir une densité multimodale à  $m$  modes.  $m$  est fixé à l'avance, ( $m$  peut être le nombre de modes de  $f$  par exemple). On choisit ensuite une famille de densité paramétrique  $\mathcal{F} = \{f_\theta; \theta \in \Theta \subset \mathbb{R}^q\}$ . La densité d'importance cherchée peut alors s'écrire  $\tilde{f}(x) = \sum_{i=1}^m \lambda_i f_{\theta_i}(x)$  ( $\sum_{i=1}^m \lambda_i = 1$ ); on notera  $\tilde{f}(\cdot, \lambda)$  au lieu de  $\tilde{f}(\cdot)$  et le rapport de vraisemblance  $w(\cdot, \alpha)$  au lieu de  $w(\cdot)$ . La démarche est la suivante :

1. Étape 1 : Sélection d'une densité d'importance initiale en sélectionnant  $m$ , et  $\theta_i$  et en initialisant  $\lambda$ .
2. Étape 2 : Génération d'un échantillon de taille  $N$   $(X_j^{(i)})_{j=1, \dots, N}$  suivant  $f_{\theta_i}$  pour  $i = 1, \dots, m$ .

3. Étape 3 : On cherche  $\lambda$  qui minimise :

$$\hat{\delta}^2(\lambda, X, N) = \frac{\sum_{i=1}^m \lambda_i \widehat{\text{Var}}_i(w)}{(\sum_{i=1}^m \lambda_i \bar{w}_i)^2},$$

où  $\widehat{\text{Var}}_i(w) = (\sum_{j=1}^{N_i} w^2(X_j^{(i)}, \lambda) / (N_i - \bar{w}_i^2))$ ,  $\bar{w}_i = \sum_{j=1}^{N_i} w(X_j^{(i)}, \lambda) / N_i$  et  $N_i = [\lambda_i N]$ .

4. On recommence l'étape 3 avec le nouveau paramètre  $\lambda$  et on stoppe l'algorithme lorsque l'erreur commise entre le nouveau  $\hat{\delta}^2$  et le précédent  $\hat{\delta}^2$  devient inférieure à la précision  $\varepsilon$  voulue.

L'efficacité de cette technique repose sur un bon choix initial de  $\lambda$ .  $N$  ne doit pas être trop grand. Le principal défaut de cette technique est le choix par avance du nombre de modes.

### Mélanges défensifs

T. Hesterberg [Hes 95] propose une approche par mélange qu'il nomme « mélange défensif », la densité d'importance  $\tilde{f}_\lambda$  est le mélange de la densité initiale  $f$  et d'une autre densité  $\tilde{f}_0(x)$  :

$$\tilde{f}_\lambda(x) = \lambda f(x) + (1 - \lambda) \tilde{f}_0(x), \quad 0 \leq \lambda \leq 1. \quad (2.2.21)$$

Le mélange défensif a le principal avantage de rendre le rapport de vraisemblance borné :

$$w(x) = \frac{f(x)}{\tilde{f}_\lambda} \leq \frac{1}{\lambda} \quad \forall x \quad (2.2.22)$$

Nous avons vu (cf. § 2.2.2.1) qu'un rapport de vraisemblance borné apporte une certaine robustesse à la méthode du tirage d'importance. La variance asymptotique de  $\hat{I}'$  est majorée (2.2.17) :

$$\text{Var}(\hat{I}') \leq \frac{1}{\lambda} \text{Var}(\hat{I}^{MC}).$$

L'utilisation du mélange défensif est un bon compromis car on est sûr de ne pas augmenter la variance de l'estimateur considérablement sans pour autant rechercher de manière intensive une densité d'importance appropriée. Idéalement,  $\tilde{f}$  devrait s'approcher de  $f^*$  quand  $f^*(x) \geq f(x)$ ,  $\tilde{f}_0$  devrait permettre de tirer dans les queues de  $g$ .

Il reste quand même la détermination de  $\lambda$ , un  $\lambda$  élevé est approprié si on sait que  $\tilde{f}_0$  s'approche de  $f^*$  ( $\tilde{f}_0$  a été déterminé auparavant) et un  $\lambda$  est approprié dans le cas contraire. On peut aussi choisir le  $\lambda$  qui minimise la variance de l'estimateur grâce à une étude préliminaire.

### 2.3 Tirage d'importance adaptatif

On parle de tirage d'importance adaptatif lorsque l'on estime  $I$  à partir d'une suite de densités d'importance  $(\tilde{f}_j)_{j=0, \dots, k-1}$  construite afin d'approcher  $f^*$  quand  $k \rightarrow +\infty$ . Le calcul

de  $\tilde{f}_j, j = 1, \dots, k-1$  prend en compte les réalisations de variables aléatoires indépendantes et identiquement distribuées selon  $\tilde{f}_{j-1}, j = 1, \dots, k-1$ ,  $\tilde{f}_0$  représente la densité initiale fixée arbitrairement. On peut aussi estimer  $I$  à chaque étape et ainsi établir un critère d'arrêt de l'algorithme.

Nous noterons dans la suite  $\mathcal{F}_{k-1}$  la tribu engendrée par  $X_1^{(0)}, \dots, X_{n_1}^{(0)}, \dots, X_1^{(k-1)}, \dots, X_{n_{k-1}}^{(k-1)}$ , où  $X_i^{(j)} \sim \tilde{f}_j, j = 0, \dots, k-1, i = 1 \dots, n_j$ .

### 2.3.1 Construction d'un estimateur adaptatif

Nous aborderons ici plusieurs manières de construire un estimateur de  $I$  à partir d'une suite de densités :

1. estimateur par tirage d'importance classique avec le dernier terme  $\tilde{f}_{k-1}$  de la suite en tant que densité d'importance et des réalisations indépendantes  $X_i^{(k-1)}, X_i^{(k-1)} \sim \tilde{f}_{k-1}$  (§ 2.3.1.1) ;
2. estimateur construit en utilisant tous les termes de la suite  $(\tilde{f}_j)_{j=0, \dots, k-1}$  et des réalisations  $X_i^{(j)}, j = 0, \dots, k-1, i = 1 \dots, n_{j+1}$  (termes non identiquement distribués) (§ 2.3.1.2) ;
3. estimateur par tirage d'importance classique avec une densité d'importance  $\check{f}_{k-1}$  construite en utilisant un mélange de tous les termes de la suite, et des réalisations indépendantes de  $\check{X}_i^{(k-1)}, \check{X}_i^{(k-1)} \sim \check{f}_{k-1}$  (§ 2.3.1.3).

Nous allons voir dans les paragraphes qui suivent les différents estimateurs, en supposant connue la méthode de construction de cette suite de densités. Nous verrons aussi sous quelles conditions ces estimateurs conservent de bonnes propriétés. Ensuite, dans la section 2.3.2, nous nous intéresserons à la construction de la suite de densités  $(\tilde{f}_j)$ .

#### 2.3.1.1 Estimateur simple à k étapes

##### . Estimateur

Pour construire cet estimateur en  $k$  étapes, les  $k-1$  premières étapes permettent de déterminer de façon adaptative la densité  $\tilde{f}_{k-1}$  et la dernière étape permet d'estimer par tirage d'importance simple l'intégrale  $I$  :

- étape 1 : réalisation de  $n_1$  tirages suivant  $\tilde{f}_0$  pour déterminer  $\tilde{f}_1$
- étape 2 : réalisation de  $n_2$  tirages suivant  $\tilde{f}_1$  pour déterminer  $\tilde{f}_2$
- ...
- étape  $k$  : réalisation de  $n_k$  tirages suivant  $\tilde{f}_{k-1}$  pour calculer  $\hat{I}_k$

L'estimateur de  $I$  a alors la forme suivante :

$$\hat{I}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} g(X_i^{(k-1)}) \frac{f(X_i^{(k-1)})}{\tilde{f}_{k-1}(X_i^{(k-1)})} \text{ avec } X^{(k-1)} \sim \tilde{f}_{k-1}. \quad (2.3.1)$$

Cet estimateur est un estimateur sans biais de  $I$ . La loi des grands nombres assure la convergence presque sûre de  $\hat{I}_k$  vers  $I$  car les termes de l'estimateur sont indépendants et identiquement distribués suivant  $\tilde{f}_{k-1}$ .

### . Variance

Pour alléger les équations, nous noterons  $\mathcal{W}_{k-1}(x) = g(x) \frac{f(x)}{\tilde{f}_{k-1}(x)} = g(x)w_{k-1}(x)$ .

La variance de  $\hat{I}_k$  est :

$$\text{Var}(\hat{I}_k) = \frac{1}{n_k} \left\{ \mathbb{E} \left[ \mathbb{E}_{\tilde{f}_{k-1}} \left[ \mathcal{W}_{k-1}^2(X^{(k-1)}) \right] \right] - I^2 \right\} \quad (2.3.2)$$

Un estimateur de cette variance est donné par :

$$\widehat{\text{Var}}(\hat{I}_k) = \frac{1}{n_k(n_k - 1)} \sum_{i=1}^{n_k} \left( g^2(X_i^{(k-1)})w_{k-1}^2(X_i^{(k-1)}) - \hat{I}_k^2 \right). \quad (2.3.3)$$

### . Exemples

$k = 1$  : Réalisation de  $n_1$  tirages suivant  $\tilde{f}_0$  pour calculer  $\hat{I}_1$  :  $\hat{I}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathcal{W}_0(X_i^{(0)})$ .

$k = 2$  : Réalisation de  $n_1$  tirages suivant  $\tilde{f}_0$  pour déterminer  $\tilde{f}_1$  et réalisation de  $n_2$  tirages suivant  $\tilde{f}_1$  pour calculer  $\hat{I}_2$  :  $\hat{I}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} \mathcal{W}_1(X_i^{(1)})$ ,  $X_i^{(i)} \sim \tilde{f}_1$ .

### . Remarque

Nous pouvons aussi estimer  $I$  par une variante de l'estimateur  $\hat{I}_k$  (cf. §2.1.8) :

$$\hat{I}'_k = \frac{1}{\sum_{i=1}^{n_k} w_{k-1}(X_i^{(k-1)})} \sum_{i=1}^{n_k} g(X_i^{(k-1)})w_{k-1}(X_i^{(k-1)}) \text{ avec } X^{(k-1)} \sim \tilde{f}_{k-1}.$$

#### 2.3.1.2 Estimateur avec recyclage à k étapes

### . Estimateur

Cet estimateur se construit à partir des densités intermédiaires  $(\tilde{f}_i)_{i=0, \dots, (k-1)}$ . Cela permet de « recycler » les tirages que l'on a effectués pour adapter les paramètres. Nous pouvons estimer  $I$  par :

$$\hat{I}_k = \sum_{j=1}^k c_j \hat{I}_j, \text{ avec } \sum_{i=1}^k c_j = 1.$$

En prenant  $c_j = n_j/n^{(k)}$ ,  $n^{(k)} = \sum_{j=1}^k n_j$ , nous obtenons l'estimateur suivant :

$$\hat{I}_k = \frac{1}{n^{(k)}} \sum_{j=1}^k \sum_{i=1}^{n_j} \mathcal{W}_{j-1} \left( X_i^{(j-1)} \right), \quad X^{(j-1)} \sim \tilde{f}_{j-1}. \quad (2.3.4)$$

Si  $n_j = 1 \forall j$ , on ne fait qu'un seul tirage à chaque étape, et l'estimateur est donné par :

$$\hat{I}_k = \frac{1}{k} \sum_{j=1}^k \mathcal{W}_{j-1} \left( X_i^{(j-1)} \right).$$

Etudions maintenant l'estimateur  $\hat{I}_k$  (équation (2.3.4)) :

### . Biais

$\hat{I}_k$  est un estimateur sans biais de  $I$ .

*Démonstration :*

*En remarquant :*

$$\mathbb{E}_{\tilde{f}_{j-1}} \left[ \frac{g(X^{(j-1)})f(X^{(j-1)})}{\tilde{f}_{j-1}(X^{(j-1)})} \mid \mathcal{F}_{j-1} \right] = I \quad \forall j = 1, \dots, k. \quad (2.3.5)$$

*Nous obtenons :*

$$\begin{aligned} \mathbb{E} \left[ \hat{I}_k \right] &= \mathbb{E} \left[ \mathbb{E} \left( \frac{1}{n^{(k)}} \sum_{j=1}^k \sum_{i=1}^{n_j} \mathcal{W}_{j-1}(X_i^{(j-1)}) \mid \mathcal{F}_{j-1} \right) \right] \\ &= \frac{1}{n^{(k)}} \sum_{j=1}^k \sum_{i=1}^{n_j} \mathbb{E} \left[ \mathbb{E} \left( \mathcal{W}_{j-1}(X_i^{(j-1)}) \mid \mathcal{F}_{j-1} \right) \right] \\ &= \frac{1}{n^{(k)}} \sum_{j=1}^k \sum_{i=1}^{n_j} \mathbb{E} [I] = I \end{aligned}$$

■

### . Variance

La variance de  $\hat{I}_k$  s'exprime par :  $\text{Var} \left( \hat{I}_k \right) = \frac{1}{n^{(k)^2}} \sum_{j=1}^k n_j \text{Var} \left( \mathcal{W}_{j-1} \left( X_1^{(j-1)} \right) \right)$ .

*Démonstration :*

*L'espérance de l'estimateur  $\hat{I}_k$  conditionnellement à  $\mathcal{F}_{k-1}$  est donnée par :*

$$\mathbb{E} \left[ \hat{I}_k \mid \mathcal{F}_{k-1} \right] = \frac{1}{n^{(k)}} \left\{ \sum_{j=1}^{k-1} \sum_{i=1}^{n_j} \mathcal{W}_{j-1} \left( X_i^{(j-1)} \right) + n_k I \right\}. \quad (2.3.6)$$

*La variance de l'estimateur  $\hat{I}_k$  conditionnellement à  $\mathcal{F}_{k-1}$  s'écrit :*

$$\text{Var} \left[ \hat{I}_k \mid \mathcal{F}_{k-1} \right] = \frac{n_k}{n^{(k)^2}} \cdot \text{Var} \left( \mathcal{W}_{k-1}(X_1^{(k-1)}) \right). \quad (2.3.7)$$

Et ainsi :

$$\begin{aligned}
\text{Var}(\hat{I}_k) &= \mathbb{E} \left[ \text{Var} \left( \hat{I}_k | \mathcal{F}_{k-1} \right) \right] + \text{Var} \left( \mathbb{E} \left[ \hat{I}_k | \mathcal{F}_{k-1} \right] \right) \\
&= \mathbb{E} \left[ \frac{n_k}{n^{(k)^2} \cdot \text{Var} \left( \mathcal{W}_{k-1}(X_1^{(k-1)}) \right) \right] + \text{Var} \left( \frac{1}{n^{(k)}} \sum_{j=1}^{k-1} \sum_{i=1}^{n_j} \mathcal{W}_{j-1} \left( X_i^{(j-1)} \right) \right) \\
&= \frac{n_k}{n^{(k)^2} \cdot \text{Var} \left( \mathcal{W}_{k-1}(X_1^{(k-1)}) \right) + \left( \frac{n^{(k-1)}}{n^{(k)}} \right)^2 \text{Var} \left( \frac{1}{n^{(k-1)}} \sum_{j=1}^{k-1} \sum_{i=1}^{n_j} \mathcal{W}_{j-1} \left( X_i^{(j-1)} \right) \right) \\
&= \frac{1}{n^{(k)^2} \sum_{j=1}^k n_j \text{Var} \left( \mathcal{W}_{j-1} \left( X_1^{(j-1)} \right) \right) \\
&= \frac{1}{n^{(k)^2} \sum_{j=1}^k n_j \left\{ \mathbb{E} \left[ \mathbb{E}_{\tilde{f}_{j-1}} \left[ \mathcal{W}_{j-1}^2 \left( X^{(j-1)} \right) \right] \right] - I^2 \right\} \tag{2.3.8}
\end{aligned}$$

■

## • Convergences

Nous supposons que  $n^{(k)} \rightarrow +\infty$  quand  $k \rightarrow +\infty$ .

1)  $\hat{I}_k$  converge dans  $L^2$  vers  $I$  quand  $k \rightarrow +\infty$  sous la condition (d'après (2.3.8)) :

$$\frac{1}{n^{(k)^2} \sum_{j=1}^k n_j \text{Var} \left[ \mathcal{W}_{j-1} \left( X_1^{(j-1)} \right) \right] \xrightarrow{k \rightarrow +\infty} 0.$$

2)  $\hat{I}_k$  converge presque sûrement vers  $I$  quand  $k \rightarrow +\infty$  sous la condition :

$$\sum_{j=1}^{\infty} \frac{n_j \text{Var} \left[ \mathcal{W}_{j-1} \left( X^{(j-1)} \right) \right]}{(n^{(j)})^2} < +\infty.$$

Nous retrouverons cette condition au paragraphe 2.3.2.1, nous pouvons nous y référer pour voir quand celle-ci est satisfaite.

*Démonstration :*

Tout d'abord, nous introduisons le théorème suivant :

### **Théorème 2.3.1** [Fel 71]

Soit  $(X_k)$  une suite de v.a.  $\mathcal{F}_{k-1}$ -adaptée  $\mathbb{E}[X_k | \mathcal{F}_{k-1}] = 0, \forall k$ .

Si  $b_1 < b_2 < \dots < b_k < \dots \rightarrow +\infty$  et si  $\sum_{j=1}^{+\infty} \mathbb{E}[X_j^2] / b_j^2 < +\infty$ , alors :

$$\frac{X_1 + X_2 + \dots + X_k}{b_k} \xrightarrow{k \rightarrow +\infty} 0 \text{ p.s.}$$

Posons  $\psi_j = \sum_{i=1}^{n_j} (\mathcal{W}_{j-1}(X_i^{(j-1)}) - I)$  et  $b_k = n^{(k)}$  et vérifions les conditions du théorème :

$$\begin{aligned}
(i) \quad \mathbb{E}[\psi_j | \mathcal{F}_{j-1}] &= \sum_{i=1}^{n_j} \left\{ \underbrace{\mathbb{E} \left[ \mathcal{W}_{j-1} \left( X_i^{(j-1)} \right) \mid \mathcal{F}_{j-1} \right]}_{I \text{ d'après (2.3.5)}} - \mathbb{E}[I | \mathcal{F}_{j-1}] \right\} \\
&= 0 \quad \forall j.
\end{aligned}$$

$$\begin{aligned}
(ii) \quad & \sum_{j=1}^{+\infty} \frac{\mathbb{E}[\psi_j^2]}{b_j^2} < +\infty \\
\iff & \sum_{j=1}^{+\infty} \frac{\text{Var}(\psi_j)}{b_j^2} < +\infty \\
& \text{car } \mathbb{E}[\psi_j] = 0 \\
\iff & \sum_{j=1}^{+\infty} \frac{n_j \text{Var}(\mathcal{W}_{j-1}(X^{(j-1)}))}{(n^{(j)})^2} < +\infty \\
& \text{car } \text{Var}(\psi_j) = \mathbb{E}[\text{Var}(\psi_j|\mathcal{F}_{j-1})] + \underbrace{\text{Var}(\mathbb{E}[\psi_j|\mathcal{F}_{j-1}])}_0
\end{aligned}$$

Par le théorème, nous obtenons :

$$\begin{aligned}
& \frac{\psi_1 + \dots + \psi_k}{b_k} \xrightarrow[k \rightarrow +\infty]{} 0 \text{ p.s.} \\
\iff & \frac{1}{n^{(k)}} \sum_{j=1}^k \left( \sum_{i=1}^{n_j} \mathcal{W}_{j-1}(X_i^{(j-1)}) - I \right) \xrightarrow[k \rightarrow +\infty]{} 0 \text{ p.s.} \\
\iff & \frac{1}{n^{(k)}} \sum_{j=1}^k \sum_{i=1}^{n_j} \mathcal{W}_{j-1}(X_i^{(j-1)}) - I \xrightarrow[k \rightarrow +\infty]{} 0 \text{ p.s.} \\
\iff & \hat{I}_k \xrightarrow[k \rightarrow +\infty]{} I \text{ p.s.}
\end{aligned}$$

■

## • Remarques

1. Comparons les conditions de convergence presque sûre et  $L^2$  dans un cas simple :  $n_j = n \forall j$ .

$$(\text{cond 1}) : \sum_{j=1}^{\infty} \frac{\text{Var}[\mathcal{W}_{j-1}(X^{(j-1)})]}{j^2} < +\infty,$$

$$(\text{cond 2}) : \frac{1}{k^2} \sum_{j=1}^k \text{Var}[\mathcal{W}_{j-1}(X_1^{(j-1)})] \xrightarrow[k \rightarrow +\infty]{} 0$$

$$\text{Si } \text{Var}[\mathcal{W}_{j-1}(X_1^{(j-1)})] \sim j^\alpha :$$

$$(\text{cond1}) \iff \sum_{j=1}^{\infty} j^{\alpha-2} < +\infty \Rightarrow \alpha < 1$$

$$(\text{cond2}) \iff \frac{k^{1+\alpha}}{k^2} \rightarrow 0 \Rightarrow \alpha < 1$$

2. Comme pour l'estimateur simple à  $k$  étapes, nous pouvons utiliser une variante pour estimer  $I$  :

$$\hat{I}'_k = \frac{1}{\sum_{j=1}^k \sum_{i=1}^{n_j} w_{j-1}(X_i^{(j-1)})} \sum_{j=1}^k \sum_{i=1}^{n_j} \mathcal{W}_{j-1}(X_i^{(j-1)}), \quad X^{(j-1)} \sim \tilde{f}_{j-1}.$$

### 2.3.1.3 Estimateur mélange

Cet estimateur s'écrit comme un tirage d'importance simple avec une densité d'importance qui est un mélange des densités intermédiaires :

$$\hat{I}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} g(X_i) \frac{f(X_i)}{\check{f}_{k-1}(X_i)} \text{ avec } X_i \sim \check{f}_{k-1}, \quad (2.3.9)$$

$$\text{où } \check{f}_{k-1}(x) = \sum_{i=0}^{k-1} \alpha_i^{(k-1)} \tilde{f}_i(x), \quad \sum_{i=0}^{k-1} \alpha_i^{(k-1)} = 1.$$

Etudions les propriétés de cet estimateur :

•  $\hat{I}_k$  est un estimateur sans biais de  $I$  car nous avons :

$$\mathbb{E}_{\check{f}_{k-1}} \left[ \frac{g(X)f(X)}{\check{f}_{k-1}(X)} \mid \mathcal{F}_{k-1} \right] = I \quad \forall k. \quad (2.3.10)$$

• La variance de  $\hat{I}_k$  est donnée par :

$$\text{Var}(\hat{I}_k) = \frac{1}{n_k} \left\{ \mathbb{E} \left[ \mathbb{E}_{\check{f}_{k-1}} \left[ g^2(X) \frac{f^2(X)}{\check{f}_{k-1}^2(X)} \right] \right] - I^2 \right\}. \quad (2.3.11)$$

Pour  $k = 2$  :  $\check{f}_1(x) = \alpha \tilde{f}_0(x) + (1-\alpha) \tilde{f}_1(x)$ . On se retrouve alors dans le cas de mélanges défensifs présenté au paragraphe 2.2.2.3. En supposant que la densité  $\tilde{f}_0$  est la densité originale  $f$  de  $X$ , le rapport de vraisemblance est majoré :

$$w(x) = \frac{\tilde{f}_0(x)}{\alpha \tilde{f}_0(x) + (1-\alpha) \tilde{f}_1(x)} \leq \frac{1}{\alpha}$$

Notons que si l'on dispose d'une suite de densités  $(\tilde{f}_i)_{i=1, \dots, k-1}$  et que l'on souhaite utiliser les mélanges défensifs, il est plus intéressant de choisir  $\check{f}$  telle que  $\check{f}(x) = \alpha \tilde{f}_0(x) + (1-\alpha) \tilde{f}_{k-1}(x)$ .

## 2.3.2 Recherche de la densité d'importance

Dans cette partie, nous allons aborder la construction de la suite de densités d'importance.

Nous nous plaçons ici dans le cas paramétrique. La densité d'importance  $f_\gamma$  appartient à une famille de densités choisie  $\mathcal{F} = \{f_\gamma; \gamma \in \Gamma \subset \mathbb{R}^q\}$ ,  $f_\gamma$  est continue en  $\gamma$ . Cette famille de densités doit correspondre à des variables aléatoires faciles à générer et telle que  $\text{supp}(f) \subset \text{supp}(f_\gamma)$ . Nous ne cherchons plus  $f^*$  mais  $f_{\gamma^*}$ , où  $\gamma^*$  est le  $\gamma \in \Gamma$  qui satisfait un certain critère. L'idée est de construire une suite  $(\gamma_k)$  qui converge vers  $\gamma^*$  quand  $k \rightarrow +\infty$ .

Nous commencerons par présenter au paragraphe 2.3.2.1, l'algorithme proposé par M. Oh et J. Berger [BO 92], au paragraphe 2.3.2.2, nous verrons l'algorithme de construction de  $(\gamma_k)$  basé

sur la minimisation de la variance, et au paragraphe 2.3.2.3, nous aborderons la construction de  $(\gamma_k)$  à partir de la minimisation de l'entropie croisée.

### 2.3.2.1 Optimisation sur les moments de la densité paramétrique

Supposons que le vecteur  $\gamma$  puisse être défini par des fonctions des moments de la densité d'importance choisie. Soit  $\lambda$  le premier moment de  $X$  et  $C$  la matrice de covariance de  $X$ ,  $X \sim f_\gamma$ . On considère que  $\gamma = (\lambda, C)$ . L'idée est de choisir une densité d'importance dans la famille  $\mathcal{F}$  qui a les deux mêmes premiers moments que ceux de la densité optimale.

D'après (2.1.5), nous obtenons les relations suivantes :

$$\lambda^* = \mathbb{E}_{f^*}[X] = \frac{\int x |g(x)| f(x) dx}{\int |g(y)| f(y) dy}, \quad (2.3.12)$$

$$C^* = \mathbb{E}_{f^*}[XX^t] - \mathbb{E}_{f^*}[X]\mathbb{E}_{f^*}[X]^t = \frac{\int xx^t |g(x)| f(x) dx}{\int |g(y)| f(y) dy} - \lambda^* \lambda^{*t}. \quad (2.3.13)$$

Nous construisons les suites  $(\lambda_k)$  et  $(C_k)$  à partir de la formule (2.3.12). Soient  $\lambda_0$  et  $C_0$  les valeurs initiales, les valeurs des suites  $(\lambda_k)$  et  $(C_k)$  sont définis par :

$$\lambda_k = \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} X_i^{(j-1)} \mathcal{W}_{j-1} \left( X_i^{(j-1)} \right)}{\sum_{j=1}^k \sum_{i=1}^{n_j} \mathcal{W}_{j-1} \left( X_i^{(j-1)} \right)}, \quad (2.3.14)$$

$$C_k = \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} \left( X_i^{(j-1)} \right) \left( X_i^{(j-1)} \right)^t \mathcal{W}_{j-1} \left( X_i^{(j-1)} \right)}{\sum_{j=1}^k \sum_{i=1}^{n_j} \mathcal{W}_{j-1} \left( X_i^{(j-1)} \right)} - \lambda_k \lambda_k^t, \quad (2.3.15)$$

avec  $X^{(j-1)} \sim \tilde{f}_{j-1}$  et  $\mathcal{W}_{j-1}(x) = \frac{g(x)f(x)}{\tilde{f}_{j-1}(x)}$ .

Etudions à présent la convergence presque sûre de la suite  $(\lambda_k)$ . Nous allons utiliser à nouveau le théorème 2.3.1. Pour simplifier les calculs, plaçons nous dans le cas où  $X \in \mathbb{R}$ , mais les résultats qui vont suivre peuvent être facilement généralisables au cas où  $X \in \mathbb{R}^d$ .

Posons  $\psi_j = \sum_{i=1}^{n_j} \left( X_i^{(j-1)} \mathcal{W}_{j-1} \left( X_i^{(j-1)} \right) - I \lambda^* \right)$  et  $b_k = n^{(k)}$  et vérifions les conditions du théorème :

En remarquant :

$$\begin{aligned} \mathbb{E}_{\tilde{f}_{j-1}} \left[ X^{(j-1)} \mathcal{W}_{j-1} \left( X^{(j-1)} \right) \mid \mathcal{F}_{j-1} \right] &= \mathbb{E}_f [Xg(X)] \\ &= I \mathbb{E}_{f^*}[X] \quad \text{si } g > 0 \text{ d'après (2.1.5),} \\ &= \lambda^* I, \end{aligned} \quad (2.3.16)$$

nous avons :  $\mathbb{E}[\psi_j \mid \mathcal{F}_{j-1}] = 0$ .

$$\text{De plus, } \sum_{j=1}^{+\infty} \frac{\mathbb{E}[\psi_j^2]}{b_j^2} < +\infty \iff \sum_{j=1}^{+\infty} \frac{n_j \text{Var} \left( X^{(j-1)} \mathcal{W}_{j-1} \left( X^{(j-1)} \right) \right)}{(n^{(j)})^2} < +\infty.$$

En appliquant le théorème, on obtient :

$$\frac{1}{n^{(k)}} \sum_{j=1}^k \sum_{i=1}^{n_j} X_i^{(j-1)} \mathcal{W}_{j-1} \left( X_i^{(j-1)} \right) \xrightarrow[k \rightarrow +\infty]{} \lambda^* I \text{ p.s.}$$

sous la condition  $\sum_{j=1}^{+\infty} \frac{n_j \text{Var} \left( X^{(j-1)} \mathcal{W}_{j-1} \left( X^{(j-1)} \right) \right)}{(n^{(j)})^2} < +\infty$  (1).

Nous avons vu à la section 2.3.1.2 que :

$$\frac{1}{n^{(k)}} \sum_{j=1}^k \sum_{i=1}^{n_j} \mathcal{W}_{j-1} \left( X_i^{(j-1)} \right) \xrightarrow[k \rightarrow +\infty]{} I \text{ p.s.}$$

sous la condition  $\sum_{j=1}^{+\infty} \frac{n_j \text{Var} \left( \mathcal{W}_{j-1} \left( X^{(j-1)} \right) \right)}{(n^{(j)})^2} < +\infty$  (2).

Comme  $I \neq 0$ , nous pouvons en conclure que :

$$\lambda_k \xrightarrow[k \rightarrow +\infty]{} \lambda^* \text{ p.s. sous les conditions (1) et (2)}. \quad (2.3.17)$$

De plus, en supposant que  $\gamma = \lambda$  et comme  $\tilde{f}_j (= \tilde{f}_{\lambda_j})$  appartient à une famille de densités continue en  $\lambda$ , nous avons :  $f_{\lambda_k} \xrightarrow[k \rightarrow +\infty]{} f_{\lambda^*}$  p.s. sous les conditions (1) et (2).

Nous pouvons procéder de manière similaire afin de montrer la convergence presque sûre de  $C_k$  vers  $C^*$  sous certaines conditions, quand  $k \rightarrow +\infty$ .

### Remarque sur les conditions (1) et (2) :

Ces conditions sont satisfaites dans des situations très générales. Il suffit que les variances  $\text{Var}((X^{(j-1)})^k \mathcal{W}_{j-1}(X^{(j-1)}))$  soient uniformément bornées en  $j$  pour  $k = 0, 1$  pour que les conditions soient satisfaites, car alors la convergence de la série  $\sum_{j=1}^{\infty} \frac{n_j}{(n^{(j)})^2}$  suffit pour assurer (1) et (2). Cette dernière série est toujours convergente, comme en atteste le lemme qui suit. Une condition suffisante pour que les variances  $\text{Var}((X^{(j-1)})^k \mathcal{W}_{j-1}(X^{(j-1)}))$  soient uniformément bornées en  $j$  est que  $\sup_{\gamma \in \Gamma} \{ \text{Var}(X^k g(X) f(X) / f_\gamma(X)), X \sim f_\gamma \} < \infty$ , ce qui est vrai en particulier si les rapports de vraisemblance sont bornés et si les variances de  $g(X) X^k$  sont finis.

*Lemme : Pour toute suite  $n_j$  d'entiers strictement positifs, en notant  $n^{(j)} = \sum_{k=1}^j n_k$ , nous avons :*

$$\sum_{j=1}^{\infty} \frac{n_j}{(n^{(j)})^2} < \infty$$

*Démonstration : On note que  $n_j = n^{(j)} - n^{(j-1)}$ . Comme  $n^{(j)} \geq n^{(j-1)}$ , on a pour tout  $j \geq 2$  :*

$$\frac{n_j}{(n^{(j)})^2} = \frac{n^{(j)} - n^{(j-1)}}{(n^{(j)})^2} \leq \frac{n^{(j)} - n^{(j-1)}}{n^{(j)} n^{(j-1)}} = \frac{1}{n^{(j-1)}} - \frac{1}{n^{(j)}},$$

donc

$$\sum_{j=1}^J \frac{n_j}{(n^{(j)})^2} = \frac{1}{n_1} + \sum_{j=2}^J \frac{n_j}{(n^{(j)})^2} \leq \frac{2}{n_1} - \frac{1}{n^{(J)}} \leq \frac{2}{n_1}.$$

Cela montre que la série à termes positifs  $\frac{n_j}{(n^{(j)})^2}$  est bornée donc convergente. ■

### 2.3.2.2 Minimisation de la variance

Ici, nous définissons le  $\gamma^*$  comme le  $\gamma \in \Gamma$  qui minimise la variance de l'estimateur, nous supposons l'existence et l'unicité de ce minimum :

$$\gamma^* = \operatorname{argmin}_{\gamma \in \Gamma} \left\{ \operatorname{Var} \left( g(X) \frac{f(X)}{\tilde{f}_\gamma(X)} \right), X \sim \tilde{f}_\gamma \right\}.$$

Cette expression est équivalente à :

$$\gamma^* = \operatorname{argmin}_{\gamma \in \Gamma} \left\{ \mathbb{E} \left[ g^2(X) \frac{f^2(X)}{\tilde{f}_\gamma^2(X)} \right], X \sim \tilde{f}_\gamma \right\}. \quad (2.3.18)$$

Comme l'expression de la variance théorique est inconnue, on ne peut calculer la solution optimale analytiquement, c'est pourquoi nous allons devoir résoudre le problème d'optimisation sur la variance empirique :

$$\hat{\gamma}^* = \operatorname{argmin}_{\gamma \in \Gamma} \left\{ \frac{1}{n} \sum_{i=1}^n g^2(X_i) \frac{f^2(X_i)}{\tilde{f}_\gamma^2(X_i)}, X_i \sim \tilde{f}_\gamma \right\}. \quad (2.3.19)$$

Le problème ci-dessus est difficile à résoudre car il nous faut des réalisations de variables aléatoires distribuées suivant  $\tilde{f}_\gamma$ , qui dépend de  $\gamma$  que l'on cherche à optimiser.

Pour pallier à cet obstacle, on peut réécrire (2.3.19), soit en utilisant des réalisations de variables aléatoires distribuées suivant  $f$ , soit avec des réalisations de variables aléatoires distribuées suivant  $\tilde{f}_{\gamma_0}$ , où  $\gamma_0 \in \Gamma$  est fixé arbitrairement :

$$\hat{\gamma}^* = \operatorname{argmin}_{\gamma \in \Gamma} \left\{ \frac{1}{n} \sum_{i=1}^n g^2(X_i) \frac{f(X_i)}{\tilde{f}_\gamma(X_i)}, X_i \sim f \right\}, \quad (2.3.20)$$

$$\hat{\gamma}^* = \operatorname{argmin}_{\gamma \in \Gamma} \left\{ \frac{1}{n} \sum_{i=1}^n g^2(X_i) \frac{f(X_i)}{\tilde{f}_\gamma(X_i)} \frac{f(X_i)}{\tilde{f}_{\gamma_0}(X_i)}, X_i \sim \tilde{f}_{\gamma_0} \right\}. \quad (2.3.21)$$

### Optimisation par Newton Raphson (NR) :

Dans cette partie, nous proposons une autre construction des suites  $(\lambda_k)$  et  $(C_k)$  ( $\gamma_k = (\lambda_k, C_k)$ ) à partir de l'algorithme de Newton Raphson. On se place dans le cas où  $X$ , de loi de densité  $\tilde{f}_\gamma$ , est une variables aléatoire réelle.  $\lambda \in \mathbb{R}$  et  $C = \sigma^2 \in R$  sont respectivement la moyenne et la variance de  $X$ .

Optimisation de  $\lambda$ ,  $\sigma$  fixé :

$$\text{Nous cherchons à minimiser : } F(\lambda) = \mathbb{E}_{\tilde{f}_\gamma} \left[ \frac{g^2(X)f^2(X)}{\tilde{f}_\gamma^2(X)} \right]$$

$$\begin{aligned} v(\lambda) &= F'(\lambda) = \frac{\partial}{\partial \lambda} \mathbb{E}_{\tilde{f}_\gamma} \left[ g^2(X) \frac{f^2(X)}{\tilde{f}_\gamma^2(X)} \right] \\ &\approx \frac{1}{N} \sum_{i=1}^N \frac{\partial}{\partial \lambda} \left( g^2(X_i) \frac{f^2(X_i)}{\tilde{f}_\gamma(X_i)} \right), \quad X_i \sim \tilde{f}_\gamma^2 \\ &\approx \frac{1}{N} \sum_{i=1}^N g^2(X_i) f^2(X_i) \frac{-\frac{\partial}{\partial \lambda} \tilde{f}_\gamma(X_i)}{\tilde{f}_\gamma^3(X_i)} \end{aligned}$$

Ainsi

$$v(\lambda) = 0 \iff \sum_{i=1}^N g^2(X_i) f^2(X_i) \frac{\frac{\partial}{\partial \lambda} \tilde{f}_\gamma(X_i)}{\tilde{f}_\gamma^3(X_i)} = 0$$

Soit la fonction :

$$\begin{aligned} u(\lambda) &= \sum_{i=1}^N g^2(X_i) f^2(X_i) \frac{\frac{\partial}{\partial \lambda} \tilde{f}_\gamma(X_i)}{\tilde{f}_\gamma^3(X_i)} \\ u'(\lambda) &= \sum_{i=1}^N g^2(X_i) f^2(X_i) \frac{\partial}{\partial \lambda} \left( \frac{\frac{\partial}{\partial \lambda} \tilde{f}_\gamma(X_i)}{\tilde{f}_\gamma^3(X_i)} \right) \\ &= \sum_{i=1}^N g^2(X_i) f^2(X_i) \frac{\tilde{f}_\gamma(X_i) \frac{\partial^2}{\partial \lambda^2} \tilde{f}_\gamma(X_i) - 3 \left( \frac{\partial}{\partial \lambda} \tilde{f}_\gamma(X_i) \right)^2}{\tilde{f}_\gamma^4(X_i)} \end{aligned}$$

Par l'algorithme NR,  $\lambda_{k+1} = \lambda_k - \frac{u(\lambda_k)}{u'(\lambda_k)}$  :

$$\lambda_{k+1} = \lambda_k - \frac{\sum_{i=1}^N g^2(X_i) f^2(X_i) \frac{\frac{\partial}{\partial \lambda_k} \tilde{f}_k(X_i)}{\tilde{f}_k^3(X_i)}}{\sum_{i=1}^N g^2(X_i) f^2(X_i) \frac{\tilde{f}_k(X_i) \frac{\partial^2}{\partial \lambda_k^2} \tilde{f}_k(X_i) - 3 \left( \frac{\partial}{\partial \lambda_k} \tilde{f}_k(X_i) \right)^2}{\tilde{f}_k^4(X_i)}}$$

avec  $\lambda_0$  quelconque.

Remarquons que dans le cas où  $\tilde{f}_\gamma$  est la densité d'une loi normale de moyenne  $\lambda$  et de variance  $\sigma^2$ , nous avons :

$$\frac{\partial}{\partial \lambda} \tilde{f}_\gamma(x) = \frac{\lambda(x - \lambda)}{\sigma^2} \tilde{f}_\gamma(x)$$

Nous pouvons réécrire  $v$  :

$$v(\lambda) = 0 \iff \sum_{i=1}^N g^2(X_i) f^2(X_i) \frac{\lambda(X_i - \lambda)}{\tilde{f}_\gamma(X_i)} = 0$$

$$\begin{aligned}
u(\lambda) &= \sum_{i=1}^N g^2(X_i) f^2(X_i) \frac{\lambda(X_i - \lambda)}{\tilde{f}_\gamma(X_i)} \\
u'(\lambda) &= \sum_{i=1}^N g^2(X_i) f^2(X_i) \frac{\partial}{\partial \lambda} \left( \frac{\lambda(X_i - \lambda)}{\tilde{f}_\gamma(X_i)} \right) \\
&= \sum_{i=1}^N g^2(X_i) f^2(X_i) \frac{\sigma^2(X_i - 2\lambda) - \lambda^2(X_i - \lambda)^2}{\sigma^2 \tilde{f}_\gamma(X_i)}
\end{aligned}$$

Si  $\tilde{f}_\gamma$  est une gaussienne de moyenne  $\lambda$  et de variance  $\sigma^2$ , alors la suite des  $(\lambda_k)$  qui converge vers  $\lambda^*$  est donnée par :

$$\begin{aligned}
\lambda_{k+1} &= \lambda_k - \frac{\sum_{i=1}^N g^2(X_i) f^2(X_i) \frac{\lambda_k(X_i - \lambda_k)}{\tilde{f}_k(X_i)}}{\sum_{i=1}^N g^2(X_i) f^2(X_i) \frac{\sigma^2(X_i - 2\lambda_k) - \lambda_k^2(X_i - \lambda_k)^2}{\sigma^2 \tilde{f}_k(X_i)}} \\
&= \lambda_k - \sigma^2 \frac{\sum_{i=1}^N \pi_k(X_i) \lambda_k(X_i - \lambda_k)}{\sum_{i=1}^N \pi_k(X_i) (\sigma^2(X_i - 2\lambda_k) - \lambda_k^2(X_i - \lambda_k)^2)} \text{ avec } \pi_k(x) \triangleq \frac{g^2(x) f^2(x)}{\tilde{f}_k(x)}.
\end{aligned}$$

*Optimisation simultanée de  $\lambda$  et  $\sigma$  :*

On cherche à minimiser la fonction suivante :

$$F(\lambda, \sigma^2) = \mathbb{E}_{\tilde{f}} \left[ g^2(X) \frac{f^2(X)}{\tilde{f}_\gamma^2(X)} \right]$$

On introduit alors  $v$ , le jacobien de cette fonction  $F$  :

$$\begin{aligned}
v(\lambda, \sigma^2) &= \begin{pmatrix} \frac{\partial}{\partial \lambda} \mathbb{E}_{\tilde{f}} \left[ g^2(X) \frac{f^2(X)}{\tilde{f}_\gamma^2(X)} \right] \\ \frac{\partial}{\partial \sigma^2} \mathbb{E}_{\tilde{f}} \left[ g^2(X) \frac{f^2(X)}{\tilde{f}_\gamma^2(X)} \right] \end{pmatrix} \\
&\approx \begin{pmatrix} \frac{1}{N} \sum_{i=1}^N g^2(X_i) f^2(X_i) \frac{-2 \frac{\partial}{\partial \lambda} \tilde{f}_\gamma(X_i)}{\tilde{f}_\gamma^3(X_i)} \\ \frac{1}{N} \sum_{i=1}^N g^2(X_i) f^2(X_i) \frac{-2 \frac{\partial}{\partial \sigma^2} \tilde{f}_\gamma(X_i)}{\tilde{f}_\gamma^3(X_i)} \end{pmatrix}
\end{aligned}$$

Dans la suite, nous nous plaçons dans le cas où  $\tilde{f}_\gamma$  est une gaussienne de moyenne  $\lambda$  et de variance  $\sigma^2$ , nous avons la relation :

$$\frac{\partial}{\partial \sigma^2} \tilde{f}_\gamma(x) = ((x - \lambda)^2 - \sigma^2) \frac{\tilde{f}_\gamma(x)}{2\sigma^4}$$

Et ainsi :

$$v(\lambda, \sigma^2) = 0 \iff \phi(\lambda, \sigma^2) \triangleq \begin{pmatrix} \sum_{i=1}^N g^2(X_i) f^2(X_i) \frac{\lambda(X_i - \lambda)}{\sigma^2 \tilde{f}_\gamma^2(X_i)} \\ \sum_{i=1}^N g^2(X_i) f^2(X_i) \frac{(X_i - \lambda)^2 - \sigma^2}{\sigma^4 \tilde{f}_\gamma^2(X_i)} \end{pmatrix} = 0$$

En appliquant l'algorithme de NR, nous obtenons :

$$\begin{aligned} \begin{pmatrix} \lambda_k \\ \sigma_k^2 \end{pmatrix} &= \begin{pmatrix} \lambda_k \\ \sigma_k^2 \end{pmatrix} - (J_k)^{-1} \phi(\lambda_k, \sigma_k^2). \\ J_k &= \begin{pmatrix} \sum_{i=1}^N g^2(X_i) f^2(X_i) \frac{\partial}{\partial \lambda_k} \frac{\lambda_k(X_i - \lambda_k)}{\sigma_k^2 \tilde{f}_\gamma^2(X_i)} & \sum_{i=1}^N g^2(X_i) f^2(X_i) \frac{\partial}{\partial \sigma_k^2} \frac{\lambda_k(X_i - \lambda_k)}{\sigma_k^2 \tilde{f}_\gamma^2(X_i)} \\ \sum_{i=1}^N g^2(X_i) f^2(X_i) \frac{\partial}{\partial \lambda_k} \frac{(X_i - \lambda_k)^2 - \sigma_k^2}{\sigma_k^4 \tilde{f}_\gamma^2(X_i)} & \sum_{i=1}^N g^2(X_i) f^2(X_i) \frac{\partial}{\partial \sigma_k^2} \frac{(X_i - \lambda_k)^2 - \sigma_k^2}{\sigma_k^4 \tilde{f}_\gamma^2(X_i)} \end{pmatrix} \\ &= \dots \\ &= \begin{pmatrix} \sum_{i=1}^N w_k^2(X_i) \frac{\sigma_k^2(X_i - 2\lambda_k) - 2\lambda_k^2(X_i - \lambda_k)^2}{\sigma_k^4} & \sum_{i=1}^N w_k^2(X_i) \frac{-\sigma_k^2 \lambda_k(X_i - \lambda_k) - (X_i - \lambda_k)^2 + \sigma_k^2}{\sigma_k^6} \\ \sum_{i=1}^N 2(\lambda_k - X_i) w_k^2(X_i) \frac{\lambda_k(X_i - \lambda_k)^2 - \lambda_k \sigma_k^2 + \sigma_k^2}{\sigma_k^6} & - \sum_{i=1}^N w_k^2(X_i) \frac{(X_i - \lambda_k)^4}{\sigma_k^8} \end{pmatrix} \end{aligned}$$

Un autre possibilité pour résoudre le problème d'optimisation (2.3.19) serait d'utiliser l'algorithme de Robbins-Monro [BDM 01]. Cet algorithme stochastique est bien adapté pour le problème de minimisation (2.3.18) où intervient une espérance. Cependant, d'une part, il converge trop lentement pour nos applications, d'autre part, l'algorithme NR dans le cas de densités gaussiennes (où les expressions des dérivées sont analytiques) est très efficaces pour nos applications.

### 2.3.2.3 Minimisation de l'entropie croisée

Pour estimer le paramètre  $\gamma^*$ , on peut utiliser une alternative à la minimisation de la variance qui est la minimisation de l'entropie croisée de Kullback-Leibler [RK 04]. L'entropie croisée mesure la « distance » entre deux densités de probabilité  $f$  et  $\tilde{f}$ , elle s'écrit :

$$\mathcal{D}(f, \tilde{f}) = \mathbb{E}_f \left[ \ln \frac{f(X)}{\tilde{f}(X)} \right] = \int f(x) \ln f(x) dx - \int f(x) \ln \tilde{f}(x) dx. \quad (2.3.22)$$

Nous supposons que  $\mathcal{D}(f^*, \tilde{f}_\gamma)$  atteint son minimum en  $\gamma = \gamma^* \in \Gamma$ .

$$\begin{aligned} \gamma^* &= \operatorname{argmin}_{\gamma \in \Gamma} \left\{ \mathbb{E} \left[ \ln \frac{f^*(X)}{\tilde{f}_\gamma(X)} \right], X \sim f^* \right\}, \\ \iff \gamma^* &= \operatorname{argmax}_{\gamma \in \Gamma} \left\{ \mathbb{E} \left[ \ln \tilde{f}_\gamma(X) \right], X \sim f^* \right\}, \\ \iff \gamma^* &= \operatorname{argmax}_{\gamma \in \Gamma} \left\{ \mathbb{E} \left[ g(X) \ln \tilde{f}_\gamma(X) \right], X \sim f \right\}. \end{aligned} \quad (2.3.23)$$

On ne peut calculer la solution analytique, ainsi :

$$\hat{\gamma}^* = \operatorname{argmax}_{\gamma \in \Gamma} \left\{ \frac{1}{n} \sum_{i=1}^n g(X_i) \ln \tilde{f}_\gamma(X_i), X_i \sim f \right\}. \quad (2.3.24)$$

Remarquons que l'on peut aussi écrire :

$$\hat{\gamma}^* = \operatorname{argmax}_{\gamma \in \Gamma} \left\{ \frac{1}{n} \sum_{i=1}^n g(X_i) \frac{f(X_i)}{\tilde{f}_{\gamma_0}(X_i)} \ln \tilde{f}_{\gamma}(X_i), X \sim \tilde{f}_{\gamma_0} \right\}. \quad (2.3.25)$$

La méthode d'entropie croisée et de minimisation de la variance ont toutes les deux le même objectif, estimer le  $\gamma \in \Gamma$  optimal. Il serait intéressant de comparer leur solution dans un cas particulier. Prenons une fonction  $g$  qui est une indicatrice (contexte de fiabilité des structures) :  $g(x) = \mathbf{1}_{\{G(x) \leq 0\}}(x)$ . La formule (2.3.23) est équivalente à :

$$\gamma_{ec}^* = \operatorname{argmin}_{\gamma \in \Gamma} \left\{ \mathbb{E} \left[ g(X) \ln \frac{f(X)}{\tilde{f}_{\gamma}(X)} \right], X \sim f \right\},$$

et rappelons l'expression de  $\gamma^*$  qui minimise la variance :

$$\gamma_{mv}^* = \operatorname{argmin}_{\gamma \in \Gamma} \left\{ \mathbb{E} \left[ g(X) \frac{f(X)}{\tilde{f}_{\gamma}(X)} \right], X \sim f \right\} \text{ car } g^2 = g.$$

Notons la similitude de ces deux expressions.

### 2.3.3 Critère d'optimisation de $n_k$

Le but général est en fait, pour un entier  $n$  fixé, de déterminer la suite  $(n_1, \dots, n_k)$  avec  $\sum_{i=1}^k n_i = n$  qui minimise la variance de  $\hat{I}_k$ . On cherche  $p_1, p_2, \dots, p_k \in [0, 1]$  avec  $\sum_{j=1}^k p_j = 1$  tels que notre estimateur  $\hat{I}_k$  avec  $(n_1, n_2, \dots, n_k) = ([Np_1], [Np_2], \dots, [Np_k])$  nous donne la plus petite variance empirique.

Dans la suite, nous allons nous placer dans le cas  $k = 2$ . Rappelons que l'estimation de  $I$  se fait alors en deux temps, le premier servant à construire une densité d'importance  $\tilde{f}_1$  (à partir de  $\tilde{f}_0$ ), puis le deuxième à estimer  $I$ .

#### 2.3.3.1 Comportement asymptotique de $\lambda_1$

Commençons par étudier la convergence en loi de  $\lambda_1$  (cf. § 2.3.2.1 pour sa construction).

Par le théorème central limite, nous obtenons :

$$\left( \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} (X_i^{(0)} \mathcal{W}_0(X_i^{(0)}) - \lambda^* I), \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} (\mathcal{W}_0(X_i^{(0)}) - I) \right) \xrightarrow{n_1 \rightarrow +\infty} (Z_N, Z_D) \sim \mathcal{N}(0, \Sigma^2),$$

$$\Sigma^2 = \begin{pmatrix} \operatorname{Var}(Z_N) & \operatorname{Cov}(Z_N, Z_D) \\ \operatorname{Cov}(Z_N, Z_D) & \operatorname{Var}(Z_D) \end{pmatrix}$$

$$\begin{aligned} \operatorname{Var}(Z_N) &= \operatorname{Var}_{\tilde{f}_0} \left[ X \mathcal{W}_0(X^{(0)}) \right] \\ &= I^2 \left( \int \frac{x^2 f^{*2}(x)}{\tilde{f}_0(x)} dx - \lambda^{*2} \right) \end{aligned}$$

$$\begin{aligned}\text{Var}(Z_D) &= \text{Var}_{\tilde{f}_0} [\mathcal{W}_0(X^{(0)})] \\ &= I^2 \left( \int \frac{f^{*2}(x)}{\tilde{f}_0(x)} dx - 1 \right)\end{aligned}$$

$$\begin{aligned}\text{Cov}(Z_N, Z_D) &= \mathbb{E}_{\tilde{f}_0} [Z_N Z_D] \\ &= n_1 I^2 \left( \int \frac{x f^{*2}(x)}{\tilde{f}_0(x)} dx - \lambda^* \right)\end{aligned}$$

**Proposition 2.3.1 :**

$$\sqrt{n_1} (\lambda_1 - \lambda^*) \xrightarrow{\mathcal{L}} \frac{(Z_N - \lambda^* Z_D)}{I} \text{ quand } n_1 \rightarrow +\infty.$$

*Preuve :*

Soient  $Z_N^{(n_1)} = \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} (X_i^{(0)} \mathcal{W}_0(X_i^{(0)}) - \lambda^* I)$  et  $Z_D^{(n_1)} = \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} (\mathcal{W}_0(X_i^{(0)}) - I)$ . On a alors :

$$\lambda_1 - \lambda^* = \frac{\lambda^* I + \frac{Z_N^{(n_1)}}{\sqrt{n_1}}}{I + \frac{Z_D^{(n_1)}}{\sqrt{n_1}}} - \lambda^*.$$

$$\text{Soit } Q_{n_1} = \sqrt{n_1} (\lambda_1 - \lambda^*) = \frac{(Z_N^{(n_1)} - \lambda^* Z_D^{(n_1)})}{I + \frac{Z_D^{(n_1)}}{\sqrt{n_1}}}, \text{ et } \tilde{Q}_{n_1} = \frac{(Z_N^{(n_1)} - \lambda^* Z_D^{(n_1)})}{I}.$$

$\tilde{Q}_{n_1}$  converge en loi vers  $\frac{Z_N - \lambda^* Z_D}{I}$  quand  $n_1 \rightarrow +\infty$ .

Si nous montrons la convergence en probabilité de  $Q_{n_1}$  vers  $\tilde{Q}_{n_1}$  quand  $n_1 \rightarrow +\infty$ , nous aurons alors le résultat de la proposition (2.3.1).

$$P(|Q_{n_1} - \tilde{Q}_{n_1}| > \delta) = P(|Q_{n_1} - \tilde{Q}_{n_1}| > \delta \cap |Z_D^{(n_1)}| < M) \quad (1)$$

$$+ P(|Q_{n_1} - \tilde{Q}_{n_1}| > \delta \cap |Z_D^{(n_1)}| > M) \quad (2).$$

$$(2) < P(Z_D^{(n_1)} > M) \xrightarrow{n_1 \rightarrow +\infty} P(Z_D > M) \underset{M \rightarrow +\infty}{=} 0.$$

$$\begin{aligned}(1) &= P\left(\left| \frac{(Z_N^{(n_1)} - \lambda^* Z_D^{(n_1)})(Z_D^{(n_1)}/\sqrt{n_1})}{I(I + Z_D^{(n_1)}/\sqrt{n_1})} \right| > \delta \cap |Z_D^{(n_1)}| < M\right) \\ &< P\left(|Z_N^{(n_1)} - \lambda^* Z_D^{(n_1)}| \frac{1}{\sqrt{n_1}} \frac{M}{I(I - M/\sqrt{n_1})} > \delta\right) \\ &\xrightarrow{n_1 \rightarrow +\infty} P(|Z_N - \lambda^* Z_D| > \delta) \leq P(Z_D > M) \underset{M \rightarrow +\infty}{=} 0.\end{aligned}$$

Et nous avons alors montré que :

$$P(|Q_{n_1} - \tilde{Q}_{n_1}| > \delta) \xrightarrow{n_1 \rightarrow +\infty} 0 \quad \forall \delta.$$

**Cas particulier :**

On suppose que  $f^*$  et  $\tilde{f}_0$  sont respectivement les densités des gaussiennes  $\mathcal{N}(\lambda^*, \sigma^{*2})$  et  $\mathcal{N}(\lambda_0, \sigma_0^2)$ . Nous avons les résultats suivants :

$$\begin{aligned} \int \frac{x^2 f_*^2(x)}{f_0(x)} dx &= \frac{\sigma_0^2}{\sqrt{\sigma^{*2} (2\sigma_0^2 - \sigma^{*2})}} \exp\left(\frac{(\lambda_0 - \lambda_*)^2}{2\sigma_0^2 - \sigma^{*2}}\right) \left(\text{Var}_{f_{\mathcal{N}}} [X] + (\mathbb{E}_{f_{\mathcal{N}}} [X])^2\right) \\ &= \frac{\sigma_0^2 \left(\sigma_0^2 \sigma^{*2} (2\sigma_0^2 - \sigma^{*2}) + (2\lambda_* \sigma_0^2 - \lambda_0 \sigma^{*2})^2\right)}{\sqrt{\sigma^{*2} (2\sigma_0^2 - \sigma^{*2})^5}} \exp\left(\frac{(\lambda_0 - \lambda_*)^2}{2\sigma_0^2 - \sigma^{*2}}\right) \end{aligned}$$

où la fonction  $f_{\mathcal{N}}$  représente la densité de la loi normale  $\mathcal{N}\left(\frac{2\lambda_* \sigma_0^2 - \lambda_0 \sigma^{*2}}{2\sigma_0^2 - \sigma^{*2}}, \frac{\sigma_0^2 \sigma^{*2}}{2\sigma_0^2 - \sigma^{*2}}\right)$ .

$$\begin{aligned} \int \frac{x f_*^2(x)}{f_0(x)} dx &= \frac{\sigma_0^2}{\sqrt{\sigma^{*2} (2\sigma_0^2 - \sigma^{*2})}} \exp\left(\frac{(\lambda_0 - \lambda_*)^2}{2\sigma_0^2 - \sigma^{*2}}\right) \mathbb{E}_{f_{\mathcal{N}}} [X] \\ &= \frac{\sigma_0^2}{\sqrt{\sigma^{*2} (2\sigma_0^2 - \sigma^{*2})}} \exp\left(\frac{(\lambda_0 - \lambda_*)^2}{2\sigma_0^2 - \sigma^{*2}}\right) \left(\frac{2\lambda_* \sigma_0^2 - \lambda_0 \sigma^{*2}}{2\sigma_0^2 - \sigma^{*2}}\right) \end{aligned}$$

Nous avons alors les expressions simplifiées suivantes :

$$\text{Var}(Z_N) = I^2 \left( \frac{\sigma_0^2 \left(\sigma_0^2 \sigma^{*2} (2\sigma_0^2 - \sigma^{*2}) + (2\lambda_* \sigma_0^2 - \lambda_0 \sigma^{*2})^2\right)}{\sqrt{\sigma^{*2} (2\sigma_0^2 - \sigma^{*2})^5}} \exp\left(\frac{(\lambda_0 - \lambda_*)^2}{2\sigma_0^2 - \sigma^{*2}}\right) - \lambda^{*2} \right)$$

$$\text{Var}(Z_D) = I^2 \left( \frac{\sigma_0^2}{\sqrt{\sigma^{*2} (2\sigma_0^2 - \sigma^{*2})}} \exp\left(\frac{(\lambda_0 - \lambda_*)^2}{2\sigma_0^2 - \sigma^{*2}}\right) - 1 \right)$$

$$\text{cov}(Z_N, Z_D) = n_1 I^2 \left( \frac{\sigma_0^2}{\sqrt{\sigma^{*2} (2\sigma_0^2 - \sigma^{*2})}} \exp\left(\frac{(\lambda_0 - \lambda_*)^2}{2\sigma_0^2 - \sigma^{*2}}\right) \left(\frac{2\lambda_* \sigma_0^2 - \lambda_0 \sigma^{*2}}{2\sigma_0^2 - \sigma^{*2}}\right) - \lambda^* \right)$$

### 2.3.3.2 Choix de $n_1$ et $n_2$

Dans la suite, nous nous plaçons dans le cas particulier présenté au paragraphe 2.3.3.1 précédent. Nous faisons aussi une hypothèse supplémentaire  $\gamma = \lambda$  (avant nous avons  $\gamma = (\lambda, C)$ ).  $\hat{I}_2$  représente l'estimateur simple à deux étapes de  $I$  présenté au paragraphe 2.3.1.1. D'après l'équation (2.3.2), nous avons le résultat suivant :

$$\text{Var}(\hat{I}_2) = \frac{I^2 \mathbb{E}[\Omega_1^2]}{n_2}, \quad (2.3.26)$$

$$\text{où } \Omega_1^2 = \begin{cases} \frac{\sigma_0^2}{\sqrt{\sigma^{*2}(2\sigma_0^2 - \sigma^{*2})}} \exp\left(\frac{(\lambda_1 - \lambda^*)^2}{2\sigma_0^2 - \sigma^{*2}}\right) - 1 & \text{si } 2\sigma_0^2 - \sigma^{*2} > 0 \\ +\infty & \text{sinon} \end{cases}$$

En effet, d'après l'équation (2.3.2), nous avons

$$\text{Var}(\hat{I}_2) = \frac{1}{n_2} \left\{ \mathbb{E} \left[ \mathbb{E}_{\tilde{f}_1} \left[ \mathcal{W}_1^2(X^{(1)}) \right] \right] - I^2 \right\}$$

et aussi,  $\mathbb{E}_{\tilde{f}_1} \left[ \mathcal{W}_1^2(X^{(1)}) \right] = I^2 \mathbb{E}_{f^*} \left[ \frac{f^*(X)}{\tilde{f}_1(X)} \right]$  avec :

$$\begin{aligned} \frac{f^{*2}(x)}{\tilde{f}_1(x)} &= \frac{\sqrt{2\pi\sigma_0^2}}{2\pi\sigma^{*2}} \exp\left(\frac{(x - \lambda_1)^2}{2\sigma_0^2} - \frac{(x - \lambda^*)^2}{\sigma^{*2}}\right) \\ &= \frac{\sigma_0^2}{\sqrt{\sigma^{*2}(2\sigma_0^2 - \sigma^{*2})}} f_N(x) \exp\left(\frac{(\lambda_1 - \lambda^*)^2}{2\sigma_0^2 - \sigma^{*2}}\right) \text{ si } 2\sigma_0^2 - \sigma^{*2} \geq 0, \end{aligned}$$

où la fonction  $f_N$  représente la densité de la loi normale  $\mathcal{N}\left(\frac{2\lambda_*\sigma_0^2 - \lambda_1\sigma^{*2}}{2\sigma_0^2 - \sigma^{*2}}, \frac{\sigma_0^2\sigma^{*2}}{2\sigma_0^2 - \sigma^{*2}}\right)$ .

Nous cherchons à présent une expression simplifiée de  $\text{Var}(\hat{I}_2) = \frac{I^2}{n_2} \mathbb{E}[\Omega_1^2]$ . D'après la proposition 2.3.1, nous avons :

$$\sqrt{n_1}(\lambda_1 - \lambda^*) \sim \mathcal{N}(0, \Theta^2) \text{ quand } n_1 \rightarrow +\infty,$$

$$\Theta^2 = \frac{\sigma_0^2(2\sigma_0^2 - \sigma^{*2})(\sigma_0^2\sigma^{*2} + 2\sigma_0^2 - \sigma^{*2}) + (2\lambda_*\sigma_0^2 - \lambda_0\sigma^{*2})^2}{\sqrt{\sigma^{*2}(2\sigma_0^2 - \sigma^{*2})^5}} \exp\left(\frac{(\lambda_0 - \lambda_*)^2}{2\sigma_0^2 - \sigma^{*2}}\right) - \lambda^{*2} - 1.$$

$$\mathbb{E}[\Omega_1^2] = \frac{\sigma_0^2}{\sqrt{\sigma^{*2}(2\sigma_0^2 - \sigma^{*2})}} \int \exp\left(\frac{\Theta^2 z}{n_1(2\sigma_0^2 - \sigma^{*2})}\right) f_Z(z) dz - 1 \text{ si } 2\sigma_0^2 - \sigma^{*2} \geq 0$$

où  $f_Z(z)$  est la densité d'une loi du  $\chi^2$  à 1 degré de liberté. Or :

$$\int \exp\left(\frac{\Theta^2 z}{n_1(2\sigma_0^2 - \sigma^{*2})}\right) f_Z(z) dz = \begin{cases} \left(1 - 2\frac{\Theta^2}{n_1(2\sigma_0^2 - \sigma^{*2})}\right)^{-1/2} & \text{si } \frac{\Theta^2}{n_1(2\sigma_0^2 - \sigma^{*2})} - \frac{1}{2} < 0 \\ +\infty & \text{sinon.} \end{cases}$$

D'où l'expression de la variance de  $\hat{I}_2$  :

$$\text{Var}(\hat{I}_2) = \begin{cases} \frac{I^2}{n_2} \left\{ \frac{\sigma_0^2}{\sqrt{\sigma^{*2}(2\sigma_0^2 - \sigma^{*2})}} \left(1 - 2\frac{\Theta^2}{n_1(2\sigma_0^2 - \sigma^{*2})}\right)^{-1/2} - 1 \right\} \\ \text{si } 2\sigma_0^2 - \sigma^{*2} > 0 \text{ et si } \frac{\Theta^2}{n_1(2\sigma_0^2 - \sigma^{*2})} - \frac{1}{2} < 0 \\ +\infty \text{ sinon.} \end{cases} \quad (2.3.27)$$

Regardons à présent, l'estimateur avec recyclage à deux étapes de  $I$  présenté au paragraphe 2.3.1.2, nous le noterons  $\check{I}_2$ . Nous avons le résultat suivant :

$$\text{Var}(\check{I}_2|\mathcal{F}_1) = \frac{I^2}{(n_1 + n_2)^2} (n_1\Omega_0^2 + n_2\Omega_1^2), \quad (2.3.28)$$

$$\text{où } \Omega_0^2 = \begin{cases} \frac{\sigma_0^2}{\sqrt{\sigma^{*2}(2\sigma_0^2 - \sigma^{*2})}} \exp\left(\frac{(\lambda_0 - \lambda^*)^2}{2\sigma_0^2 - \sigma^{*2}}\right) - 1 & \text{si } 2\sigma_0^2 - \sigma^{*2} > 0 \\ +\infty & \text{sinon} \end{cases}$$

En se servant des résultats précédents, et  $\sigma_1 = \sigma_0$ , nous obtenons l'expression suivante pour la variance de  $\check{I}_2$  :

$$\text{Var}(\check{I}_2) = \begin{cases} \frac{I^2}{(n_1 + n_2)^2} \left\{ n_1 \left( \frac{\sigma_0^2}{\sqrt{\sigma^{*2}(2\sigma_0^2 - \sigma^{*2})}} \exp\left(\frac{(\lambda_0 - \lambda^*)^2}{2\sigma_0^2 - \sigma^{*2}}\right) - 1 \right) \right. \\ \quad \left. + n_2 \left( \frac{\sigma_0^2}{\sqrt{\sigma^{*2}(2\sigma_0^2 - \sigma^{*2})}} \left( 1 - 2 \frac{\Theta^2}{n_1(2\sigma_0^2 - \sigma^{*2})} \right)^{-1/2} - 1 \right) \right\} \\ \quad \text{si } 2\sigma_0^2 - \sigma^{*2} > 0 \text{ et si } \frac{\Theta^2}{n_1(2\sigma_0^2 - \sigma^{*2})} - \frac{1}{2} < 0 \\ +\infty \text{ sinon.} \end{cases} \quad (2.3.29)$$

Les expressions des variances (2.3.27) et (2.3.29) nous imposent un choix de  $n_2$  grand pour l'obtention d'une petite variance, elles indiquent aussi un choix de  $n_1$  minimal en fonction des paramètres initiaux et optimaux, afin que les variances ne soient pas infinies :

$$n_1^c > \frac{2}{2\sigma_0^2 - \sigma^{*2}} \Theta^2 \quad \text{si } 2\sigma_0^2 - \sigma^{*2} > 0. \quad (2.3.30)$$

De plus, le choix de la variance est important pour l'obtention d'une variance finie,  $\sigma_0^2$  ne doit pas être « trop petit » :  $\sigma_0 > \sigma^*/\sqrt{2}$ .

Nous allons illustrer dans la section suivante l'importance des choix de  $n_1$  et  $n_2$  dans le cas où nous disposons d'un nombre fixe de simulations totales  $n$ ,  $n = n_1 + n_2$ , et l'influence du choix des paramètres initiaux sur  $n_1$  et  $n_2$ . Nous nous placerons dans le cas particulier où les calculs ont été effectués, *i.e.* le cas gaussien.

### 2.3.3.3 Illustrations

$\Theta^2$  est en fait une constante ne dépendant que des paramètres de la densité initiale  $\lambda_0$  et  $\sigma_0^2$  et des paramètres de la densité optimale *i.e.*  $\lambda_*$  et  $\sigma^*$ . La figure 2.3.1 illustre l'évolution  $\Theta^2$  en fonction du choix des paramètres optimaux, dans le cas particulier où  $\lambda_* = 20$  et  $\sigma^* = 2$ .

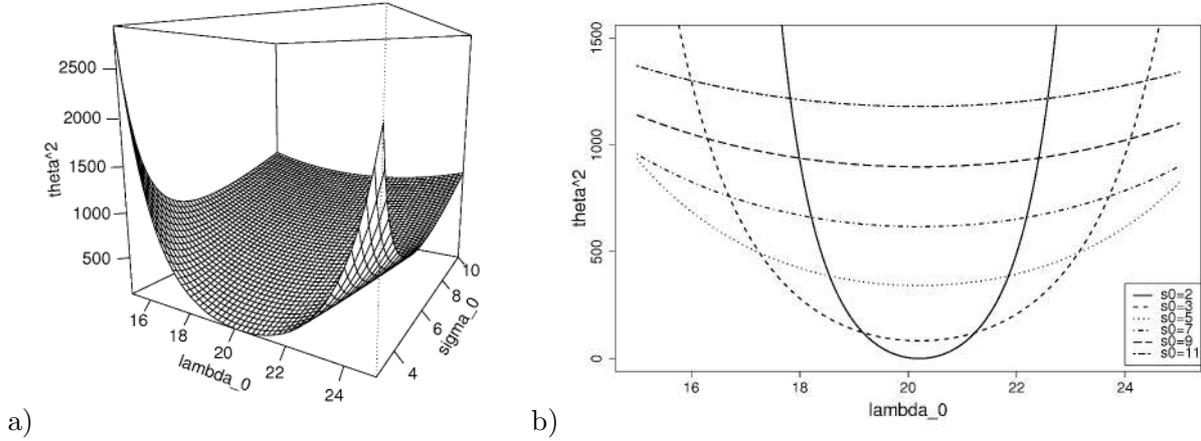


FIG. 2.3.1 – Évolution de  $\Theta^2$  en fonction des paramètres initiaux  $\lambda_0$  et  $\sigma_0$ , ( $\lambda^* = 20$  et  $\sigma^* = 2$ ).  
Figure a) :  $\Theta^2$  en fonction de  $\lambda_0$  et  $\sigma_0$ ; Figure b) :  $\Theta^2$  en fonction de  $\lambda_0$ ,  $\sigma_0$  fixe.

Intéressons nous maintenant à  $n_1^c$  défini en (2.3.30), qui nous indique le nombre  $n_1$  minimal afin de ne pas avoir une variance infinie, ses lignes de niveaux en fonction de  $\lambda_0$  et  $\sigma_0^2$ , ainsi que celles de  $\Theta^2$  sont données figure 2.3.2. Comme nous ne gérons pas l'éloignement de  $\lambda_0$  et de  $\lambda^*$ , ces graphiques nous indiquent un compromis dans le choix de  $\sigma_0$ , un trop grand  $\sigma_0$  diminue le  $n_1^c$  mais fait aussi exploser  $\Theta^2$ , en outre,  $\sigma_0$  ne doit pas être trop petit.

Intéressons nous à présent à l'évolution de  $\mathbb{E}[\Omega^2]/n_2$  en fonction de  $n_1$ , pour  $n = n_1 + n_2$  fixe.

### Illustration 1

Prenons le cas où  $\Theta^2$  est grand car  $\lambda_0$  est éloigné de  $\lambda^*$ . La figure 2.3.3 illustre le comportement de  $\mathbb{E}[\Omega^2]/n_2$  en fonction de  $n_1$ .

### Illustration 2

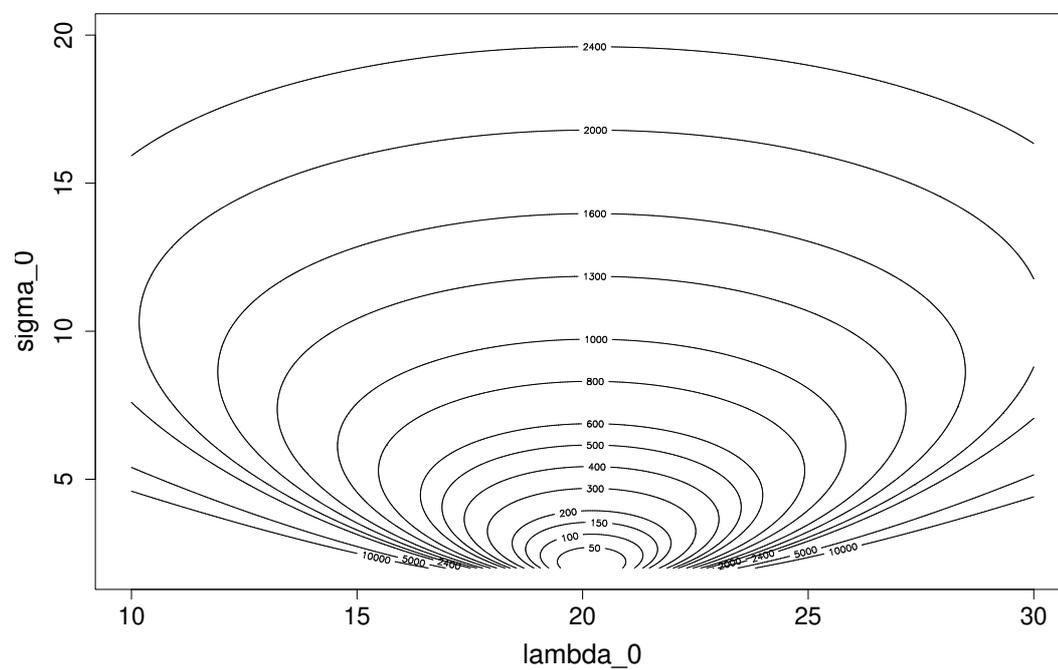
Prenons le cas où  $\Theta^2$  est grand car  $\sigma_0$  est petit. La figure 2.3.4 illustre le comportement de  $\mathbb{E}[\Omega^2]/n_2$  en fonction de  $n_1$ .

### Illustration 3

Prenons le cas où  $\Theta^2$  est petit car  $\lambda_0$  est proche de  $\lambda^*$ . La figure 2.3.5 illustre le comportement de  $\mathbb{E}[\Omega^2]/n_2$  en fonction de  $n_1$ .

En définitive, Le  $n_1$  permet de réduire la quantité  $\Theta^2$ , c'est-à-dire qu'il permet de gérer l'éloignement potentiel de  $\lambda_0$  de  $\lambda^*$ . le  $n_2$  quand à lui tente de réduire la quantité  $\mathbb{E}[\Omega^2]$ , c'est à dire la variance de l'estimateur.

a)



b)

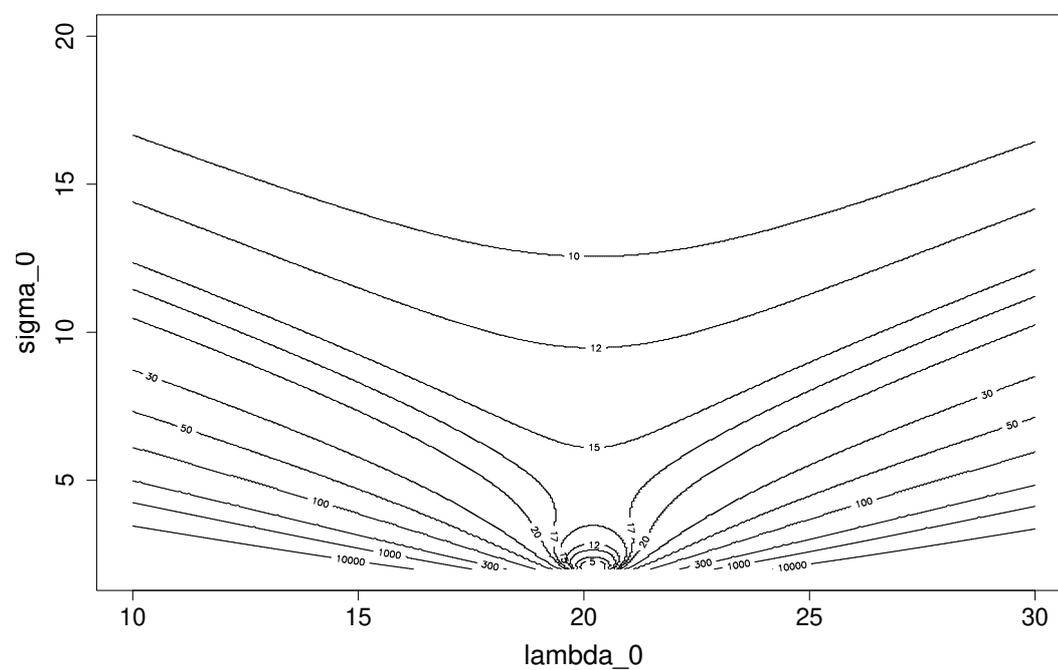


FIG. 2.3.2 – Figure a) : lignes de niveau de  $\Theta^2$  en fonction de  $\lambda_0$  et  $\sigma_0^2$ ; Figure a) : lignes de niveau de  $n_1^c$  en fonction de  $\lambda_0$  et  $\sigma_0^2$ .

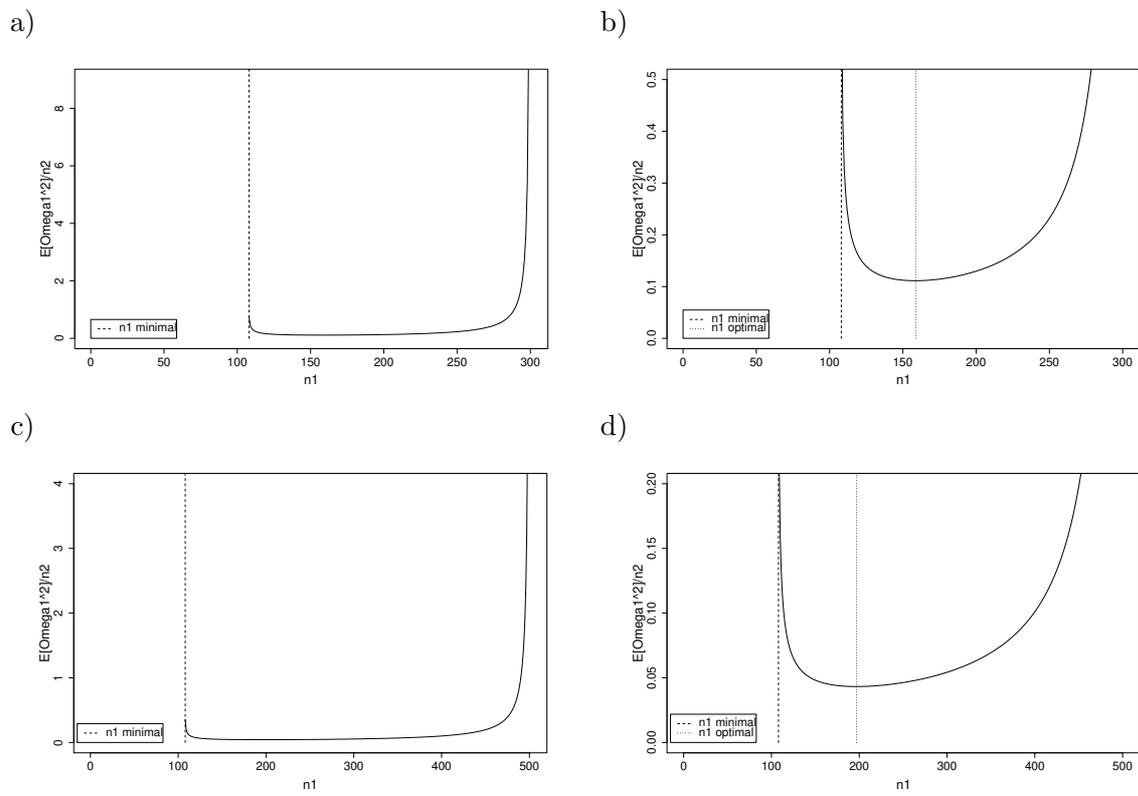


FIG. 2.3.3 – Évolution de  $\frac{E[\Omega^2]}{n_2}$  en fonction de  $n_1$ ,  $\lambda_0 = 10$ ,  $\sigma_0 = 6$ ,  $\lambda^* = 20$  et  $\sigma^* = 2$ . Figure a) :  $n = 300$  ; figure b) : zoom de a) ; figure c) :  $n = 500$  ; figure d) : zoom de c).

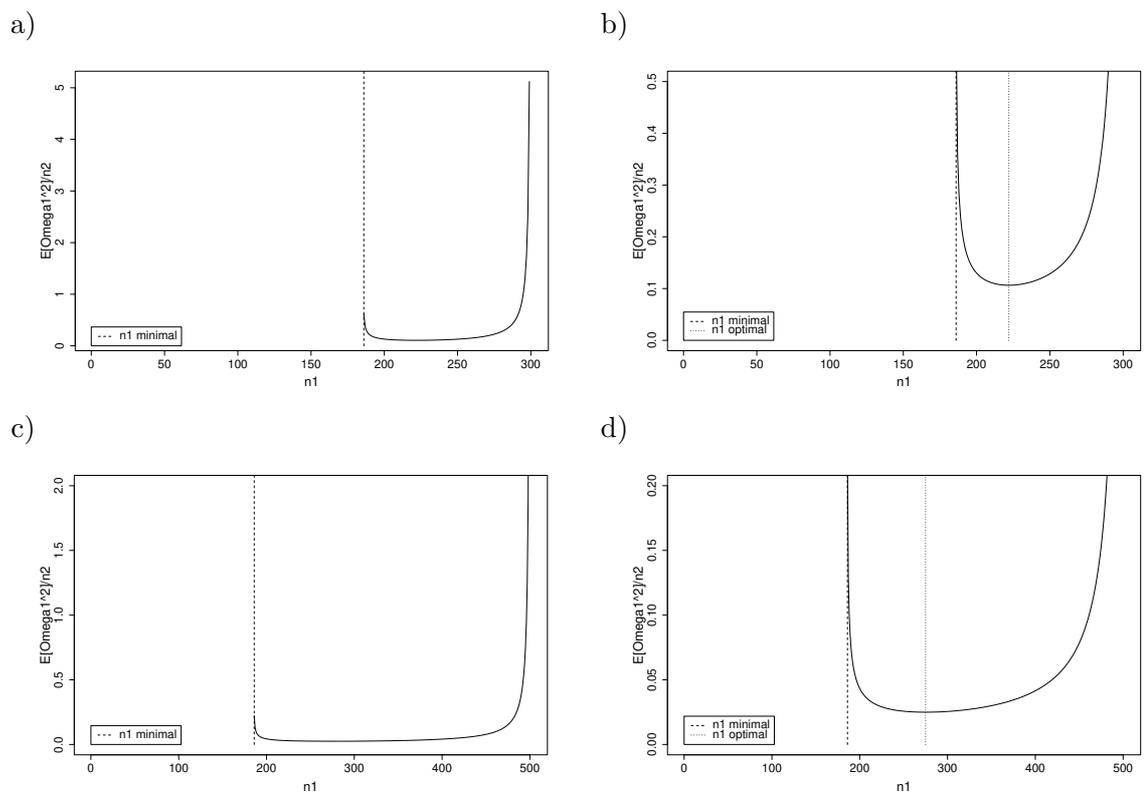


FIG. 2.3.4 – Évolution de  $\frac{E[\Omega^2]}{n_2}$  en fonction de  $n_1$ ,  $\lambda_0 = 16$ ,  $\sigma_0 = 3$ ,  $\lambda^* = 20$  et  $\sigma^* = 2$ . Figure a) :  $n = 300$  ; figure b) : zoom de a) ; figure c) :  $n = 500$  ; figure d) : zoom de c).

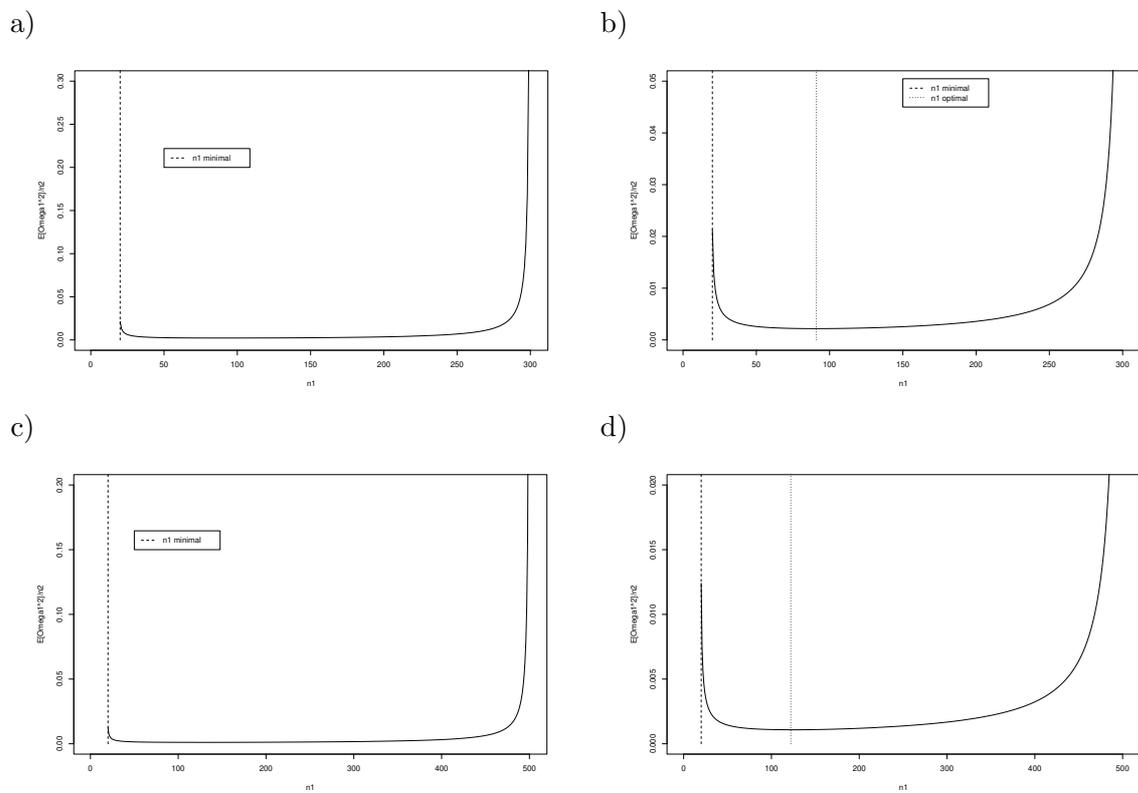


FIG. 2.3.5 – Évolution de  $\frac{\Omega^2}{n_2}$  en fonction de  $n_1$ ,  $\lambda_0 = 19$ ,  $\sigma_0 = 3$ ,  $\lambda^* = 20$  et  $\sigma^* = 2$ . Figure a) :  $n = 300$  ; figure b) : zoom de a) ; figure c) :  $n = 500$  ; figure d) : zoom de c).

## 2.4 Simulations sur des cas tests

Nous présentons dans ce chapitre quelques performances et comportements des méthodes de tirage d'importance adaptatif paramétrique qui ont été implémentées. La méthode du tirage d'importance adaptatif paramétrique a été programmée de façon générique, pour y faire varier la méthode d'estimation des paramètres des fonctions d'importance et le choix de l'estimateur final. Leur efficacité est discutée sur des fonctions exemples complètement sorties du contexte des HTR, car nous ne voulons pas optimiser une méthode sur un cas trop particulier et qui ne serait pas réutilisable dans un autre contexte. Pour cela, nous avons créé des fonctions tests qui gardent tout de même des caractéristiques particulières, dont celles d'être monotones sur deux morceaux et d'avoir une moyenne sur l'échantillon réel de  $1.e-5$ . Nous nous limitons dans cette partie à des fonction à un seul paramètre (une dimension). Les méthodes seront d'abord commentées sur des population assez nombreuses pour en dégager des comportements représentatifs. Dans un deuxième temps, nous testerons et comparerons ces méthodes sur des échantillons très réduits. Nous testerons dans la chapitre suivant l'efficacité de ces méthodes le code ATLAS ou sur des simulants du code ATLAS.

### 2.4.1 Fonctions simples utilisées pour tester les méthodes

Nous nous intéressons aux cas où la densité d'importance optimale est très décalée par rapport à la densité initiale. Dans ces conditions, un monte-carlo conditionnel seul ne donnera pas un bon résultat et le tirage d'importance prend tout son sens. Par ailleurs, nous considérerons des fonctions toujours positive.

Nous allons étudier le comportement sur plusieurs exemples simples en 1 dimension. Dans tous les cas, nous utilisons pour  $f$  la fonction normale centrée réduite de densité  $\phi$  et de fonction de répartition  $\Phi$ . Par ailleurs, nous nous plaçons dans le cas où la valeur de l'intégrale à approcher vaut  $1.10^{-5}$ .

**exemple 1 :** dans cet exemple, la fonction  $g$  est une fonction de heaviside,  $f^*$  est donc une gaussienne tronquée.

$$\left\| \begin{array}{ll} \text{si } x < \Phi^{-1}(10^{-5}) & g(x) = 1 \\ \text{sinon} & g(x) = 0 \end{array} \right. \quad (2.4.1)$$

**exemple 2 :** nous voulons que  $f^*$  soit symétrique, la fonction  $g$  valant 1 jusqu'au centre de

symétrie.  $g$ , exprimée ci dessous décroît donc très vite après ce point.

$$\left\| \begin{array}{ll} \text{si } x < \Phi^{-1}(5.10^{-6}) & g(x) = 1 \\ \text{sinon} & g(x) = \frac{\phi(x - 2 * \Phi^{-1}(5.10^{-6}))}{\phi(x)} \end{array} \right. \quad (2.4.2)$$

**exemple 3 :** nous voulons obtenir pour  $f^*$  une gaussienne excentrée par rapport à l'origine et d'écart type pas trop petit 2. Remarquons que dans ces conditions, la fonction  $g$  n'est pas partout bornée à 1.

$$\left\| g(x) = \frac{\phi(2 * (x + 4)) * (10^{-5})}{\phi(x)} \right. \quad (2.4.3)$$

Les densités des fonctions  $f^*$  pour les trois exemples sont présentés sur la figure 2.4.1

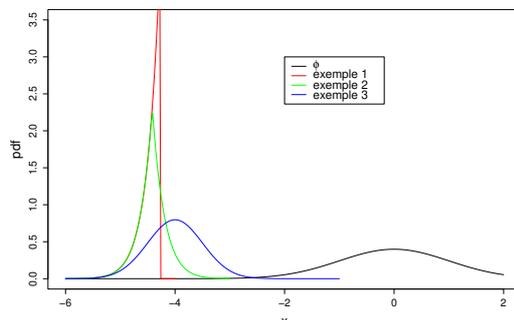


FIG. 2.4.1 – Densité d'importance optimale  $f^*$  et densité initiale  $\phi$  - exemples 1, 2 et 3

## 2.4.2 Recherche de la densité d'importance

Avant d'effectuer l'estimation finale de  $I$ , il convient de tendre le plus rapidement et le plus précisément possible vers une densité proche de la densité d'importance optimale. Lorsque le nombre d'étape est supérieur à 2, l'estimation des paramètres des fonctions d'importance doit être fait dans le but d'optimiser la réévaluation des mêmes paramètres à l'étape suivante, mais cela suit la même logique que l'estimation de  $I$ . D'ailleurs les étapes intermédiaires seront souvent utilisées pour l'estimation finale. Pour simplifier les choses, nous nous basons toujours pour ces étapes sur les méthodes d'estimation de la fonction d'importance.

### 2.4.2.1 Estimateur empirique des moments avec recyclage

Comme premiers estimateurs, nous utilisons les estimations définies par les formules (2.3.12) et (2.3.12). L'estimateur  $\lambda_k$  s'écrit

$$\lambda_k = \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} X_i^{(j-1)} \mathcal{W}_{j-1} \left( X_i^{(j-1)} \right)}{\sum_{j=1}^k \sum_{i=1}^{n_j} \mathcal{W}_{j-1} \left( X_i^{(j-1)} \right)} \quad (2.4.4)$$

et la variance estimée  $C_k$  s'écrit

$$C_k = \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} \left( X_i^{(j-1)} \right)^2 \mathcal{W}_{j-1} \left( X_i^{(j-1)} \right)}{\sum_{j=1}^k \sum_{i=1}^{n_j} \mathcal{W}_{j-1} \left( X_i^{(j-1)} \right)} \quad (2.4.5)$$

C'est à dire que chaque point se voit attribuer le poids  $\mathcal{W}_{j-1}(\cdot) = \frac{g(\cdot)f(\cdot)}{f_{j-1}(\cdot)}$ , conservé d'une étape à l'autre. La moyenne des  $X_i$  et des  $X_i^2$  pondérée est alors réalisée. Cette estimation effectue donc un recyclage des tirages.

Lors de l'utilisation du tirage d'importance adaptatif sur l'exemple 1, avec comme densité initiale  $f_0$  la densité réelle  $f$ . Si nous obtenons aucun tirage tel que  $g > 0$  (ce qui a la probabilité  $10^{-5.n}$ ), nous ne pouvons pas réévaluer la moyenne et la variance pour l'étape suivante. Dans tout ce chapitre, ce problème ne sera pas réglé. Notons tout de même que l'utilisation du tirage d'importance autour du point de conception fournit de très bons résultats mais ce n'est pas ce qui est étudié ici.

Avec les fonctions exemples 2 et 3, où nous sommes assurés d'avoir des poids non nuls à chaque point, nous obtenons des poids pour les points extrêmes de gauche très prépondérants. La conséquence de cela est que la moyenne estimée pour la fonction d'importance est située sur le(s) point(s) de gauche et la variance estimée est en général trop faible. Nous illustrons cela sur la figure 2.4.2; dans cet exemple, l'écart type estimé pour la fonction d'importance était de 0.1625, alors que l'optimum est 0.5. A plusieurs dimensions, tous les paramètres peuvent se retrouver aussi centrés sur ce(s) même(s) point(s), avec une variance faible.

Lors des étapes suivantes, le mécanisme se reproduit, la moyenne se décalant lentement vers la moyenne de la fonction d'importance. Le déplacement est à chaque fois borné par l'extrémum du tirage précédent, décalé de façon insuffisante de l'origine à cause des faibles variances successives. Nous pouvons observer ce phénomène sur les illustrations de la figure 2.4.3. Dans le cas 2, la variance estimée est faible, mais correspond en ordre de grandeur à la variance de  $f^*$ , par contre pour l'exemple 3, il faudra plusieurs étapes pour que la variance réaugmente jusqu'à la variance de  $f^*$ .

Comme cela a été indiqué, la convergence vers  $f_{\gamma^*}$  est assurée (cf. § 2.3.2.1), mais elle peut être très lente.

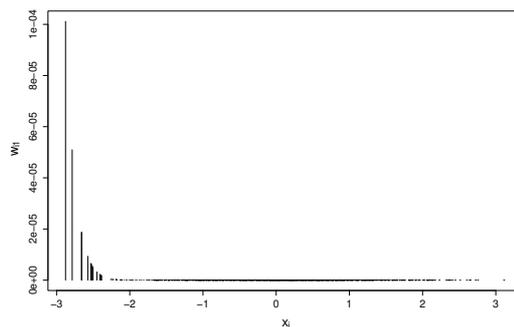


FIG. 2.4.2 – Graphe des poids des réalisations dans l'estimation des paramètres de la fonction d'importance à l'étape 1 - exemple 2

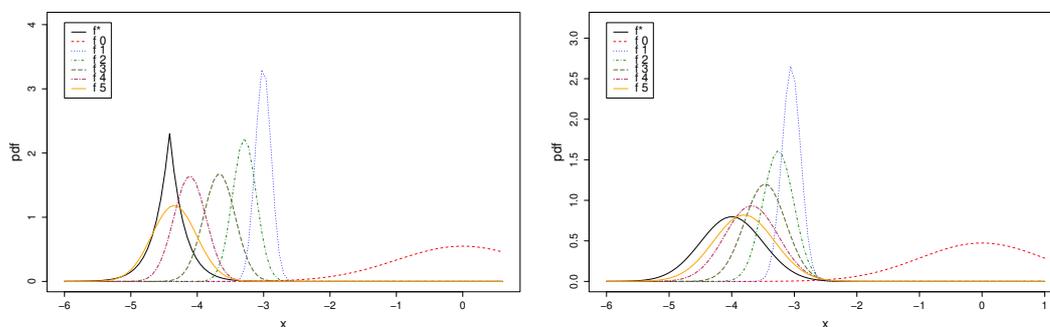


FIG. 2.4.3 – Évolution des fonctions d'importances, exemple 2 à gauche, exemple 3 à droite, exemple 2 à gauche, exemple 3 à droite

Un autre problème peut aussi survenir : sur une des étapes un poids excessif peut être dû à un tirage avec une probabilité faible (selon  $f_j$ ). Avec la méthode de recyclage utilisée, ce poids excessif peut nuire à tout l'algorithme. Dans le cas de la figure 2.4.4 le poids d'un point est très élevé. Son poids reste élevé par rapport aux tirages des étapes suivantes. Il faut plusieurs étapes pour que le point finisse par perdre sa prépondérance. De manière similaire, dans la figure 2.4.5, nous présentons les poids à l'étape 2. Comme la variance des tirages à cette étape est faible et que la fonction  $f^*$  reste non négligeable à droite de la zone probable, un tirage de faible probabilité dans cette zone à un poids élevé (alors que cela n'est pas souhaitable), ce qui renverse l'évolution vers  $f^*$  de la suite des  $f_i$  pour plusieurs étapes.

Avec cette méthode nous avons eu besoin d'environ 4 étapes à 1000 tirages pour être avec une probabilité élevée dans la zone de la meilleure fonction d'importance. Ce nombre total d'appels à la fonction  $g$  (nécessitant une simulation) est beaucoup trop élevé.

Le problème des variances intermédiaire trop faible et du risque qui en découle d'avoir des poids excessif est un problème récurrent pour lequel plusieurs pistes vont être étudiées.

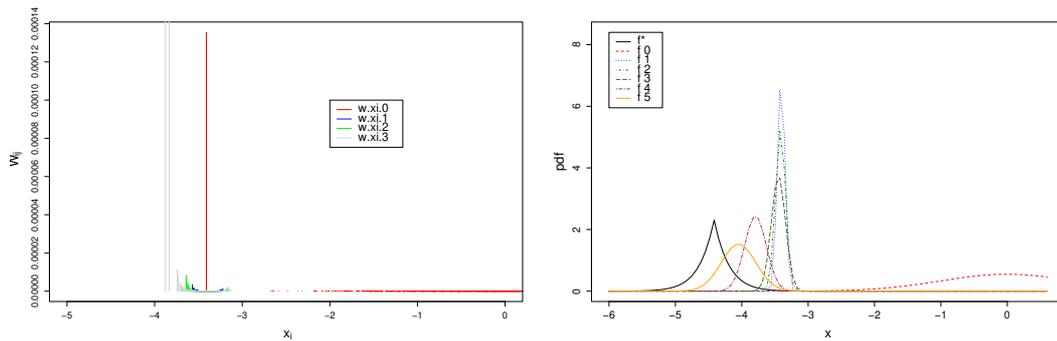


FIG. 2.4.4 – Illustration du cas d’une réalisation avec poids élevé à l’issue de la première étape. A gauche : poids des points pour les premières étapes, à droite : fonctions d’importances des différentes étapes

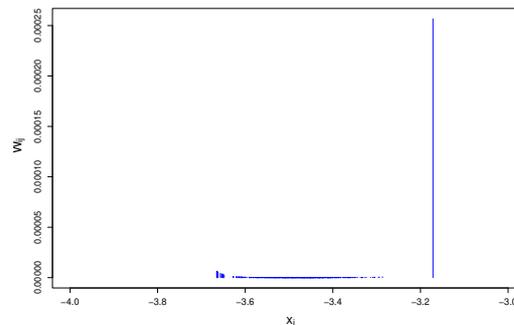


FIG. 2.4.5 – Illustration du cas d’une réalisation avec poids élevé à l’issue de la deuxième étape dans la mauvaise direction.

### 2.4.2.2 Augmentation de la variance

Si la suite des  $\lambda_k$  tend vers  $\lambda^*$  avec une vitesse limitée par la variance, nous pouvons espérer une convergence plus rapide avec une variance volontairement augmentée. Ainsi, nous effectuons les deux modifications suivantes :

- Supposant que la zone d’importance optimale sera décalée par rapport à la densité initiale, il est bénéfique d’augmenter la variance des tirages de la première étape. Par la suite, nous prendrons pour celle-ci le double de la variance de la densité initiale. Lorsque le code de simulation sera utilisé, cela pourra nécessiter de tronquer la densité initiale pour éviter de tomber dans une zone qui risque de faire échouer le code. Ce choix sur la densité initiale peu se faire à partir d’une connaissance a priori de l’influence de certains paramètres important.
- Pour les étapes suivantes, nous utilisons une variance toujours majorée par rapport à l’estimation qui en est faite. Dans notre cas, il faut éviter les estimations très

mauvaise, dues à une variance de la densité d'importance trop faible, quitte à perdre en précision dans de nombreux cas. L'avantage de cette majoration est double : d'une part, cela évite plus souvent les poids très élevés sur un unique point et d'autre part, cela permet d'aller explorer plus vite les régions d'intérêt hors des zones initialement probables.

Sur l'exemple 1, la méthode reste inefficace puisque nous avons encore dans près de 80% des cas aucune valeur de  $g$  non nulle pour 200 tirages à la première étape.

Nous présentons des résultats pour l'exemple 2 sur le graphique 2.4.6 (à gauche).

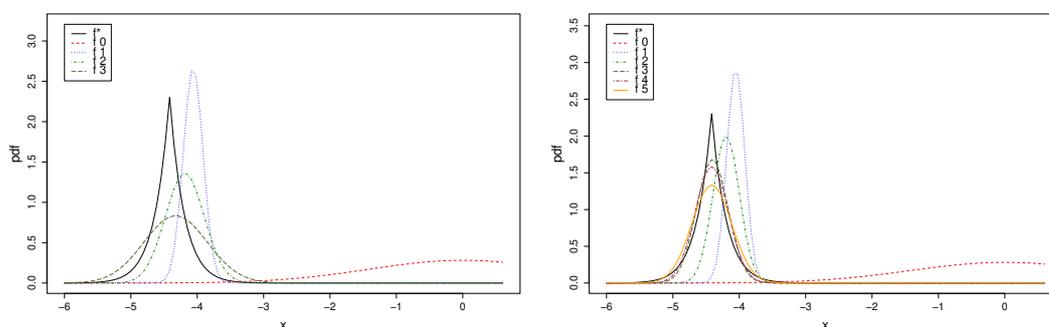


FIG. 2.4.6 – Recherche de la densité d'importance, variance doublée - exemple 2. A gauche avec 200 tirages à chaque étape et en recyclant les tirages, à droite avec 250 tirages à chaque étape et sans recycler les tirages - exemple 2

Si les résultats sont satisfaisants sur les exemples en dimension 1 et incitent à augmenter encore la variance, le fait de trop augmenter la variance va être très défavorable en dimension élevée. Si on considère, avec un écart-type double, que la moitié des essais ont un poids substantiel, les autres étant en dehors de la zone d'intérêt, en dimension  $n$ , la proportion des essais dans la zone d'intérêt devient  $0.5^n$ . Il faut donc mieux s'orienter vers une des autres solutions :

1. soit reconnaître les mauvaises estimations de la fonction d'importance pour rajouter une étape dans ces quelques cas ;
2. soit avoir un facteur sur la variance qui prends en compte l'incertitude que l'on a dessus ;
3. soit choisir pour chaque tirage une direction sur laquelle la dilatation de la variance va être appliquée.

### 2.4.2.3 Etapes intermédiaires sans recyclage

Pour éviter d'être handicapé par un poids important d'une étape précédente, comme vu précédemment, nous choisissons de n'utiliser que les tirages de la dernière étape pour estimer

les paramètres de la fonction d'importance paramétrique.

Nous continuons d'utiliser une variance augmentée, nous illustrons les résultats obtenus sur le graphique 2.4.6 (à droite) avec des étapes à 250 tirages par étape sur l'exemple 2. Nous voyons que la fonction d'importance à l'issue de l'étape 1 n'est pas encore bien estimée et a une variance trop faible. Dès la troisième étape, la fonction d'importance se stabilise autour de la fonction recherchée, avec une variance légèrement surestimée. Ce comportement standard reste satisfaisant. Remarquons que en l'absence de recyclage, nous réalisons en fait une chaîne de Markov

#### 2.4.2.4 Efficacité des estimateurs

Nous recherchons particulièrement des méthodes d'estimation efficaces, même avec peu de tirages. Nous testons maintenant la méthode de nombreuses fois sur les exemples, pour vérifier la variance et les queues de probabilité de l'estimation. Nous utilisons 3 étapes avec 200 tirages puis encore 200 tirages pour l'estimation de la probabilité finale. Dans ces conditions, nous renouvelons 5000 fois l'estimation. Pour les exemples 2 et 3, nous obtenons les histogrammes 2.4.7 et 2.4.8.

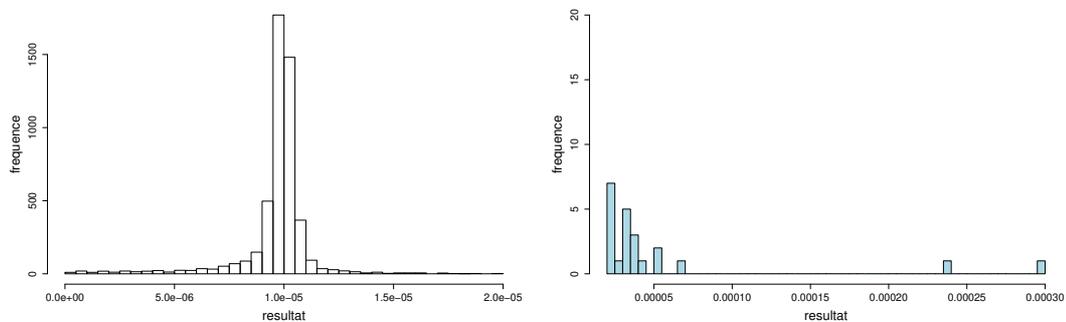


FIG. 2.4.7 – *Histogramme des résultats de l'estimation finale, sans recyclage et variance doublée - exemple 2*

Quelques résultats restent très mauvais et ce sont ces résultats mauvais que nous allons commenter. Sur la figure 2.4.9, nous voyons que les estimations trop faibles sont dues à des moyennes trop faibles pour la dernière densité d'importance. Les estimations trop fortes ont été générées à partir d'un écart type faible, mais sans que ces points ne se démarquent beaucoup de l'ensemble des points.

La figure 2.4.10 est tracée à partir des deux résultats extrêmes montrant que ces mauvaises estimations sont la suite logique d'un premier échantillon particulièrement mauvais.

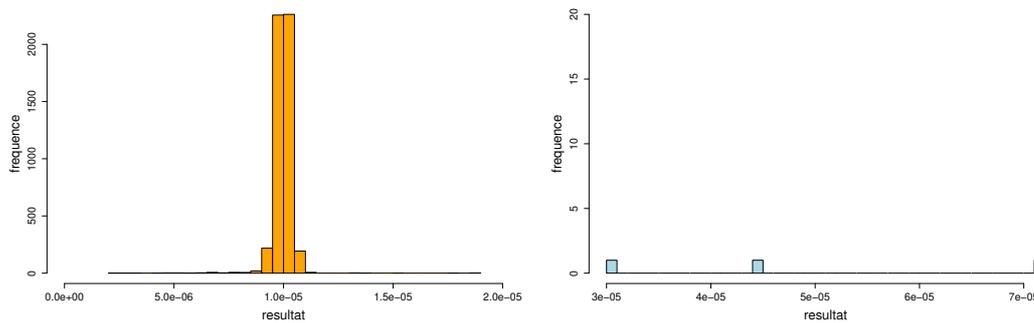


FIG. 2.4.8 – Histogramme des résultats de l'estimation finale, sans recyclage et variance doublée - exemple 3

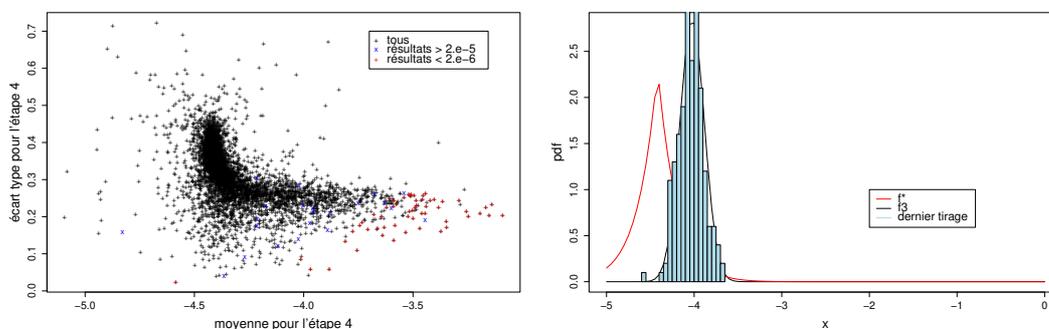


FIG. 2.4.9 – Illustration de l'influence d'une mauvaise estimation de la moyenne à l'issue de la dernière étape sur l'estimation finale, sans recyclage et variance doublée - exemple 2

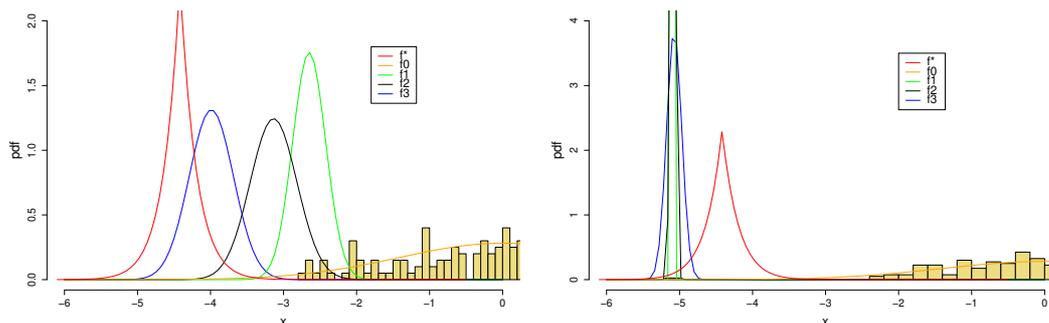


FIG. 2.4.10 – Illustration de deux suites de densités d'importance amenant à une mauvaise estimation finale - exemple 2

## 2.4.3 Performance et optimisation de $\hat{I}$ pour un nombre limité de tirages

### 2.4.3.1 Test du nombre d'étapes

Nous nous limitons à 600 appels à la fonction  $g$  au total (toutes les étapes confondues), sur l'exemple 2. Nous utilisons les estimateurs simples présentés précédemment au paragraphe 2.3.1

Nous testons les deux types d'approches des densités intermédiaires, avec et sans recyclage et les deux estimateurs pour divers découpage des 600 simulations au total.

Les résultats obtenus pour l'exemple 3 sont présentés dans le tableau 2.4.1 (résultats issus d'une série de 10000 expériences).

Tirages	Approche des paramètres	Estimateur de I	moyenne	coefficient de variation
(200, 200, 200)	Sans Recycl	Sans Recycl	$9.9 \cdot 10^{-6}$	29%
		Avec Recycl	$9.74 \cdot 10^{-6}$	56%
(200, 200, 200)	Avec Recycl	Sans Recycl	$9.854 \cdot 10^{-6}$	22%
		Avec Recycl	$9.648 \cdot 10^{-6}$	45%
(100, 100, 100, 100, 200)	Sans Recycl	Sans Recycl	$1.01 \cdot 10^{-5}$	14%
		Avec Recycl	$9.74 \cdot 10^{-6}$	37%
(100, 100, 100, 100, 200)	Avec Recycl	Sans Recycl	$9.93 \cdot 10^{-6}$	24%
((50) * 8, 200)	Sans Recycl	Sans Recycl	$9.99 \cdot 10^{-6}$	6%
((50) * 8, 200)	Avec Recycl	Sans Recycl	$9.92 \cdot 10^{-6}$	12%
((10) * 40, 200)	Sans Recycl	Sans Recycl	$9.99 \cdot 10^{-6}$	8%
((10) * 40, 200)	Avec Recycl	Sans Recycl	$9.91 \cdot 10^{-6}$	18%

TAB. 2.4.1 – Comparaison des estimateurs en faisant varier le nombre d'étapes, la méthode d'approche des paramètres et la méthode d'estimation de I sur l'exemple 3

Nous constatons que l'estimateur sans recyclage est plus performant (variance inférieure) que l'estimateur avec recyclage. Pour la recherche de la fonction d'importance, il vaut mieux faire beaucoup de petites étapes (jusqu'à un certain point sans utiliser de recyclage) que moins d'étapes avec plus de tirage dans chaque. Ne pas faire de recyclage est dans tous ces cas plus performant que de recycler. Il faut remarquer que la moyenne globale de tous ces essais est inférieure à la valeur théorique. En effet, dans un nombre non négligeable de cas, la variance des fonctions d'importance était insuffisante pour garantir la condition (2.2.14).

#### 2.4.3.2 Limitation des poids lors de la recherche de la fonction d'importance

Pour les estimations des paramètres de la fonction d'importance, nous pouvons borner les poids, c'est à dire affecter aux points de poids les plus grands une valeur plus petite. Comme cela diminue artificiellement la variance des poids, nous espérons des estimations plus

facilement convergentes, pres de la fonction d'importance optimale. Nous imposons une borne artificiellement après que les poids aient été calculés. Nous exprimons cette borne sous forme d'un nombre ou fractile  $\alpha$  de la quantité de tirages. Le poids devient :

$$w_{ij}^b = \min(w_{ij}, \text{fractile}_\alpha(W)) \quad (2.4.6)$$

où  $w_{ij}$  est exprimé par l'équation 2.4.5 et  $W$  est l'ensemble des  $w_{ij}$ .

Ces conditions évitent d'avoir un poids excessif pour l'estimation de la fonction d'importance et évitent d'avoir une fonction d'importance trop resserrée autour d'un unique point particulier. Dans tous les cas le bornage n'affecte pas l'estimateur final  $I$  et donc n'introduit pas de biais.

Remarquons que cela ne règle pas le problème des poids importants concernant l'estimation finale, i.e. des estimations mauvaises dues à un tirage en zone peu probable pour laquelle  $f^*$  beaucoup plus grande que la dernière fonction d'importance utilisée.

Nous effectuons à ce niveau, une évolution comparée, avec et sans bornes sur le poids, des moyennes successives des fonctions d'importance est présentée. Celle-ci est présentée sur la figure 2.4.11.

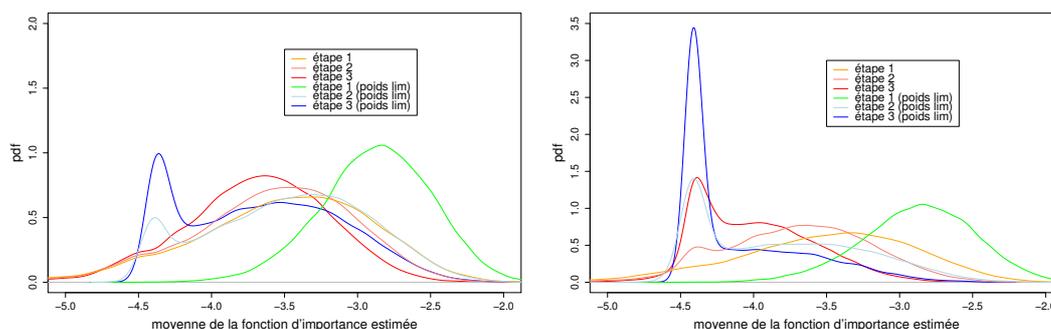


FIG. 2.4.11 – Histogramme des moyennes des densités d'importance, 100 tirages à chaque étape en bornant ou non à 4%, variance doublée - exemple 2. A gauche : avec recyclage, à droite : sans recyclage

Dans l'ensemble cela améliore les résultats de la suite des fonctions d'importances successives. Cependant, cela apporte l'inconvénient de limiter légèrement le déplacement latéral des zones d'importance lorsque les tirages sont plutôt favorables. L'effet sur l'estimateur final reste faible.

### 2.4.3.3 Utilisation du bootstrap sur un estimateur empirique

Le bootstrap peut être utilisé sur l'estimateur d'importance. d'où nous déduisons la moyenne  $\lambda_1$  et l'écart type  $\sigma_1$ . Ainsi, nous obtenons :

- la moyenne des estimations obtenus par réplication bootstrap de l'échantillon  $\lambda_{bs}$
- l'écart type de ces estimations  $sd_{bs}(\lambda)$  et  $sd_{bs}(C)$

Nous voulons que la fonction d'importance utilisée couvre le plus probablement possible toute la zone de la fonction d'importance optimale. Dans ces conditions, l'espérance et l'écart type finalement utilisés sont :

$$\begin{cases} \lambda = \lambda_{bs} \\ C = \sigma_{bs} + a * (sd_{bs}(\lambda) + sd_{bs}(C)) \end{cases} \quad (2.4.7)$$

Où  $a$  vaut 1 par défaut, mais peut être affecté à une valeur plus grande éventuellement.

Cette méthode ne présente pas de défaut qui lui soit particulier. La variance sera augmentée en fonction de l'imprécision que l'on a, donc probablement à bon escient. Les points de poids aberrants sont naturellement mis en doute par le bootstrap. Pour avoir l'estimation finale, nous pouvons utiliser les estimateurs précédents. Le tableau 2.4.2 présente les résultats obtenus sur l'exemple 3 (résultats issus d'une série de 10000 expériences) .

Tirages	Estimateur de I	moyenne	coefficient de variation
(200, 200, 200)	Sans Recycl	$9.96 \cdot 10^{-6}$	19%
(100, 100, 100, 100, 200)	Sans Recycl	$9.99 \cdot 10^{-6}$	8%
$((50) * 8, 200)$	Sans Recycl	$9.998 \cdot 10^{-6}$	4.1%

TAB. 2.4.2 – Comparaison des estimateurs en faisant varier le nombre d'étapes et méthode d'estimation de I sur l'exemple 3, la méthode d'approche des paramètres est celle utilisant le bootstrap.

### 2.4.3.4 Recyclage à partir d'une densité mixte

Soit  $f^{mix}$  la densité de probabilité artificiellement construite à l'étape  $k$  à partir des étapes précédentes par la formule :

$$f^{mix} = \frac{1}{N} \sum_{i=1}^k n_i f^i \quad (2.4.8)$$

Les échantillon tirés lors des  $k$  étapes précédentes peuvent sembler être tirées par cette fonctions, sauf que :

- la fonction  $f^i$  n'est pas indépendante des tirages des étapes précédentes ;
- les tirages réels ont été tiré en proportion fixe suivant chacune des densité constituant le mélange.

Dans ces conditions, les tirages peuvent être favorablement recyclés : tout poids très élevé impose à la densité suivante d'être élevée pour ce point et donc a pour effet de baisser tout de suite le poids correspondant.

Tant que nous ne travaillons que sur les estimations des densités d'importance, aucun biais dans l'estimation de  $I$  n'est induit.

Le tableau 2.4.3 présente les résultats obtenus pour l'exemple 3 sur les différentes méthodes de recherche des paramètres de la densité d'importance (résultats issus d'une série de 10000 expériences).

Tirages	Estimateur de $f^*$	moyenne	coefficient de variation
	simple avec recyclage(variance*2)	$9.92 \cdot 10^{-6}$	12%
	sans recyclage(variance*2)	$9.988 \cdot 10^{-6}$	6%
	limité au 4ème poids	$9.93 \cdot 10^{-6}$	17%
	bootstrap	$9.998 \cdot 10^{-6}$	4.1%
	basé sur densité mixte	$9.97 \cdot 10^{-6}$	15%

TAB. 2.4.3 – Comparaison des méthodes d'estimation des paramètres de la densité d'importance sur l'exemple 3, l'estimateur final est l'estimateur simple sans recyclage.

#### 2.4.4 Conclusion sur les estimateurs

Pour réaliser les calculs en nombre minimal, la méthode qui est apparue la meilleure est la méthode utilisant le bootstrap pour les estimation de la densité d'importance pour l'étape suivante. Pour l'estimation finale de  $I$ , l'estimation basée sur l'estimateur simple sans recyclage donne de bons résultats.

## CHAPITRE 3

### SIMULATION MULTI-NIVEAUX

La simulation multi-niveaux part du principe que l'on peut exprimer la probabilité recherchée par l'intermédiaire du produit de probabilités conditionnelles en faisant intervenir des états intermédiaires qui serviront de passerelles vers l'évènement rare.

#### 3.1 L'idée de base de la simulation multi-niveaux

L'idée de base consiste à considérer une suite décroissante de  $m$  sous-ensemble  $D_i$  emboîtés contenant l'évènement rare  $D$  :

$$D = D_m \subset D_{m-1} \subset \cdots \subset D_1 \subset D_0,$$

où  $D_0$  est l'espace d'entrée des variables. La probabilité de l'évènement  $D$  est alors donnée par :

$$\begin{aligned} P(D) &= \int_D f(x) dx \\ &= P\left(\bigcap_{i=1}^m D_i\right) \\ &= P(D_1) \prod_{i=2}^m P(D_i|D_{i-1}), \end{aligned} \tag{3.1.1}$$

où  $f$  est une densité de probabilité.

Dans un contexte de fiabilité des structures,  $D$  désigne naturellement le domaine de défaillance, et la probabilité de défaillance  $P_f$  est bien  $P(D) = P(G(x) \leq 0)$ , où  $G$  est la fonction d'état limite introduite au chapitre 1.

L'idée intuitive est d'estimer les probabilités  $p_1 = P(D_1)$  et  $p_i = P(D_i|D_{i-1})$ ,  $i = 2, \dots, m$ , séparément ; Si les sous-ensembles  $D_i$  sont choisis convenablement, les probabilités seront suffisamment grandes pour être estimées efficacement par simulation, c'est-à-dire que nous n'aurons pas besoin d'un nombre trop important de simulations pour les estimer avec précision. Par exemple, si  $p_i \sim 0.1$ ,  $i = 1, \dots, 5$ ,  $P(D) \sim 10^{-5}$  est trop faible pour être estimée efficacement par une méthode de Monte Carlo directe, alors que les probabilités intermédiaire  $p_i$  peuvent l'être car les évènements sont beaucoup plus fréquents.

Chaque probabilité  $p_i$ ,  $i = 1, \dots, m$  peut être estimée en construisant plusieurs trajectoires (une chaîne de Markov par exemple) dont les états ont pour loi asymptotique  $f(\cdot, D_{i-1})$ , où  $f(x, D_{i-1}) = \frac{f(x)\mathbf{1}_{\{D_{i-1}\}}(x)}{P(D_{i-1})}$ .

Dans la section suivante, nous allons voir comment générer de telles trajectoires par les méthodes de Monte Carlo par Chaînes de Markov. Nous verrons ensuite comment construire des estimateurs de  $p_i$ ,  $i = 1, \dots, m$ .

## 3.2 les méthodes MCMC

Les méthodes de Monte Carlo par Chaînes de Markov (MCMC) créent une longue chaîne de Markov dont les échantillons sont distribués asymptotiquement suivant une distribution requise. Pour des rappels concernant les chaînes de Markov, nous pouvons nous référer au livre de C. P. Robert et G. Casella [RC 99].

Nous présentons dans la suite deux algorithmes classiques appartenant à la famille des méthodes MCMC : l'algorithme de Metropolis-Hasting et l'algorithme de Gibbs.

### 3.2.1 Algorithme de Metropolis-Hasting

Dans cet algorithme, on se donne une distribution de transition  $x : p^*(\cdot, x)$ , cette dernière s'appelle la distribution de proposition (ou instrumentale) car c'est elle qui propose l'état suivant d'une chaîne de Markov. On se donne aussi une densité de probabilité  $\pi$ .

Pour  $i = 0, 1, 2, \dots$  :

Étant donné un état  $x_i$

- 1) Génération de  $\tilde{x}_i \sim p^*(\cdot, x_i)$ .
- 2) Calcul de  $r_i = \min \left\{ 1, \frac{\pi(\tilde{x}_i)p^*(x_i, \tilde{x}_i)}{\pi(x_i)p^*(\tilde{x}_i, x_i)} \right\}$ .
- 3) Choix de l'état suivant :

$$x_{i+1} = \begin{cases} \tilde{x}_i & \text{avec la probabilité } \min\{1, r_i\} \\ x_i & \text{avec la probabilité } 1 - \min\{1, r_i\} \end{cases}$$

Par construction, une chaîne de Markov construite par l'algorithme de Metropolis-Hasting ne se soucie pas de constantes de normalisation de  $\pi$  et de  $p^*(\cdot, x)$ , et est réversible et de loi stationnaire  $\pi$  ;

Notons que le choix de  $p^*$  est déterminant dans la performance de l'algorithme : la convergence et la vitesse de convergence dépendent du choix de la loi instrumentale.

**Remarques :** Il existe de nombreuses versions de cet algorithme :

- algorithme de Metropolis :  $p^*(x, y) = p^*(y, x)$  ;

- algorithme de Metropolis-Hasting indépendant :  $p^*(x, y) = p^*(x)$  ;
- algorithme de Metropolis-Hasting à sauts réversibles, etc ...

### 3.2.2 Gibbs

L'échantillonneur de Gibbs ne s'applique qu'à des lois multidimensionnelles de dimension  $d$ . Soient  $\pi_1, \pi_2, \dots, \pi_d$  les densités conditionnelles provenant de  $\pi$ . La démarche est la suivante :

1. initialisation de  $x_0 = (x_1^0, \dots, x_d^0)$
2. Pour  $i = 1, 2, \dots$ , et  $k = 1, \dots, d$  :  
Générer  $x_k^i \sim \pi_k(\cdot, x_{-k}^{i-1})$ .

Notons que cet algorithme accepte tous les échantillons, mais demande la connaissance des lois marginales de  $\pi$ . Il est très facile à implémenter et lorsque les lois conditionnelles sont connues, il est un meilleur choix que l'algorithme de Metropolis-Hasting.

### 3.2.3 Convergence des méthodes MCMC

Il est important de pouvoir contrôler la chaîne de Markov générée par les méthodes MCMC, c'est-à-dire, établir un critère d'arrêt qui garantit que le nombre de simulations est suffisant pour ce que l'on souhaite en faire. Plusieurs types de convergence existent :

1. convergence vers la loi stationnaire : notre échantillon est-il bien distribué selon  $\pi$  ? (c'est la convergence minimale de l'algorithme) ;
2. convergence en moyenne : A-t-on  $\frac{1}{k} \sum_{n=0}^k h(X_n)$  qui converge vers  $\int h(x) d\pi(x)$  ?
3. convergence vers un échantillon i.i.d. : à partir de quand peut-on considérer l'échantillon comme des réalisations de variables aléatoires indépendantes et identiquement distribuées selon  $\pi$  ?

Ces problèmes de convergence ont donné lieu à de nombreuses méthodes, la plus simple étant d'apprécier graphiquement la convergence de la chaîne de markov (peu fiable car la chaîne peut se bloquer dans une région pour un temps fini). Sinon, il existe plusieurs sortes de « diagnostic » : diagnostic de Geweke [Gew 92], celui de Raftery et Lewis [RL 92], celui de Gelman et Rubin [GR 92, BG 97], etc ...

### 3.2.4 Algorithme pour simuler $f(\cdot, D_i)$ , $i = 1, \dots, m - 1$

Pour simuler une chaîne de Markov admettant pour loi stationnaire  $f(\cdot, D_i)$  présentée au paragraphe 3.1 ; Au et Beck dans [AB 01] introduisent un algorithme de Métropolis légèrement modifié, il comporte une étape supplémentaire qui vérifie l'appartenance à  $D_i$  :

### Algorithme de Metropolis-Hasting modifié :

Cet algorithme correspond à celui présenté dans l'article, généralisé au cas où les composantes de  $X$  ne sont pas forcément indépendantes et la densité de proposition non symétrique.

Pour  $k = 1, 2, \dots$  :

Étant donné un état  $x_k \in D_i$ ,

1. Générer un état « candidat »  $\tilde{x}_k$  :

Soit  $\xi_k \sim p^*(\cdot, x_k)$

Poser

$$\tilde{x}_k = \begin{cases} \xi_k & \text{avec la probabilité } \min\{1, r_k\} \\ x_k & \text{avec la probabilité } 1 - \min\{1, r_k\} \end{cases}$$

$$\text{où } r_k = \frac{f(\xi_k)p^*(x_k, \xi_k)}{f(x_k)p^*(\xi_k, x_k)}$$

2. Choix de l'état suivant :

$$x_{k+1} = \begin{cases} \tilde{x}_k & \text{si } \tilde{x}_k \in D_i \\ x_k & \text{sinon} \end{cases}$$

La chaîne de Markov créée par cet algorithme admet bien comme loi stationnaire  $f(\cdot, D_i)$ .

En effet, nous avons  $p(x_{k+1}, x_k)f(x_k, D_i) = p(x_k, x_{k+1})f(x_{k+1}, D_i)$ , où  $p(\cdot, \cdot)$  désigne la densité de transition entre deux états d'une chaîne partant de  $D_i$ , alors :

$$\begin{aligned} p(x_{k+1}) &= \int p(x_{k+1}, x_k)f(x_k, D_i)dx_k \\ &= \int p(x_k, x_{k+1})f(x_{k+1}, D_i)dx_k \\ &= f(x_{k+1}, D_i) \end{aligned}$$

Une illustration de cet algorithme en dimension 2 est donnée figure 3.2.1.

## 3.3 Estimateur des probabilités conditionnelles

Pour estimer  $p_i$ ,  $i = 1, \dots, m$ , on peut tenir compte de tous les états des chaînes de Markov ou alors de seulement un nombre réduit.

### 3.3.1 Estimateur prenant en compte tous les états des chaînes

On génère plusieurs de Markov de même taille et de loi stationnaire  $f(\cdot, D_{i-1})$ . Soit  $n_i$  le nombre total de réalisations (nombre total d'états de toutes les chaînes). Soit  $n_i^c$ ,  $i = 1, \dots, m$  le nombre de chaînes de Markov simulées en parallèle, les états initiaux  $x_0^{(k)} \in D_{i-1}$ ,  $k = 1, \dots, n_i^c$  ne sont pas nécessairement tous différents. Chaque chaîne comporte alors  $n_i/n_i^c$  états.  $x_{jk}^{(i-1)}$  désigne le  $k^{\text{ième}}$  état de la  $j^{\text{ième}}$  chaîne de Markov partant du domaine  $D_{i-1}$ .

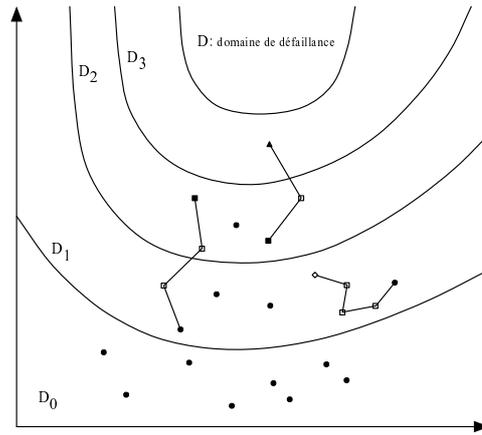


FIG. 3.2.1 – Illustration de l’algorithme de Metropolis Hasting modifié en dimension 2

Les probabilités  $p_i$  peut être estimées par une méthode de Monte Carlo directe :

$$\begin{aligned} \hat{p}_i &= \frac{1}{n_i} \sum_{j=1}^{n_i^{(c)}} \sum_{k=1}^{n_i/n_i^{(c)}} \mathbf{1}_{\{D_i\}}(x_{jk}^{(i-1)}) \\ &= \frac{1}{n_i} \sum_{k=1}^{n_i} \mathbf{1}_{\{D_i\}}(x_k), \quad x_k \sim f(\cdot, D_{i-1}). \end{aligned} \quad (3.3.1)$$

### 3.3.2 Estimateur prenant en compte quelques états des chaînes

On estimer  $p_i$  par des méthodes de splitting multi-niveaux. Nous commencerons par faire un rappel sur les méthodes de splitting.

#### 3.3.2.1 Aperçu des méthodes de splitting

Cette section n’a que l’objectif d’informer le lecteur sur les méthodes de splitting, pour une revue plus détaillée, nous pouvons nous référer à [Gar 00] ou [Lag 06], ou encore [DdFG 01]. Les méthodes de splitting font partie d’une classe de méthode efficaces pour estimer la probabilité des évènements rares. L’idée de ces méthodes repose sur l’hypothèse que les états intermédiaires seront visités plus fréquemment que l’évènement rare lui-même et ils serviront alors de passerelles vers l’évènement rare. Lorsque les probabilités conditionnelles de l’équation (3.1.1) ne sont pas connues explicitement (ce qui est presque toujours le cas), nous savons faire évoluer la trajectoire du niveau  $D_i$  au niveau  $D_{i+1}$  (processus de Markov par exemple).

Le principe de base des méthodes de splitting est alors le suivant :

1. génération simultanée de  $N_0$  trajectoires partant de l’état initial  $x_0 \in D_0$  ( $D_0 \supset D_1$ ,  $D_0$  est l’espace d’entrée des variables) ;

2. au bout d'un temps donné, certaines trajectoires ont atteint le niveau supérieur, d'autres non ; on abandonne alors les mauvaises trajectoires et on régénère des trajectoires partant des états initiaux  $x_i \in D_1, i = 1, \dots, k_1$  ; on renouvelle cette procédure jusqu'à ce que le niveau  $D$  soit atteint.

Dans l'étape b), l'arrêt au bout d'un certain temps n'est pas obligatoire, on peut définir un critère d'arrêt : par exemple, dès qu'une trajectoire atteint le niveau  $D_0$  (pour des trajectoires partant de  $D_i, i = 1 \dots, m - 1$ ), ou alors dès qu'une trajectoire est redescendue de  $d$  niveaux.

La figure 3.3.1 illustre le principe de la méthode de splitting en dimension 1.

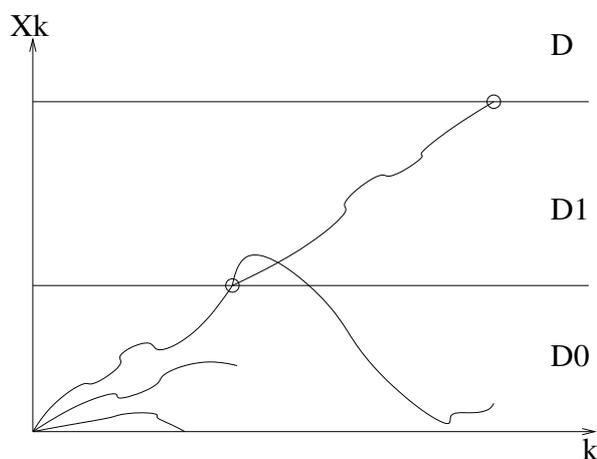


FIG. 3.3.1 – Illustration de la méthode de splitting en dimension 1

Les variantes de cette méthode concernent principalement le choix du critère d'arrêt et le nombre de trajectoires générées à chaque niveau, voici deux exemples :

**Algorithme RESTART** : [VAVA 91, VAVA 94] les trajectoires partant du niveau  $D_i$  et qui redescendent en  $D_{i-1}$  sont abandonnées sauf la dernière qui est « poursuivie » jusqu'à ce qu'elle atteigne à nouveau  $D_i$ .

**Modèle de branchement avec duplication de trajectoires** [Lag 06] : Le processus  $(x_k)$  est dupliqué en  $R_i$  processus dès qu'il entre dans un domaine  $D_i$ . La figure 3.3.2 illustre ce principe en dimension 1.

### 3.3.2.2 Estimateur par une méthode de splitting

Contrairement à l'estimateur précédent (cf. § 3.3.1), nous ne nous intéressons pas à l'évolution du processus mais au franchissement des niveaux par celui-ci. Pour estimer  $p_i$ , il est nécessaire de générer un nombre  $n_i^{(c)}$  assez important de chaînes de Markov en parallèle et

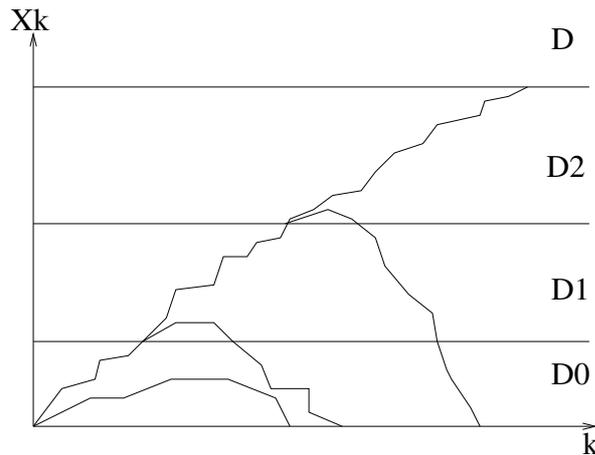


FIG. 3.3.2 – *Illustration modèle de branchement avec duplication de trajectoires en dimension 1*

partant du niveau  $D_i$ . On regarde alors le nombre de trajectoires ayant atteint le niveau supérieur pour un processus de taille fixe.

$p_1$  peut être estimée par :

$$\hat{p}_1 = \frac{R_1}{n_1^{(c)}},$$

où  $R_1$  désigne le nombre de trajectoires ayant atteint la frontière du niveau  $D_1$ .

Et  $p_i$  peut être estimée par :

$$\hat{p}_i = \frac{R_i}{n_i^{(c)} R_{i-1}} \text{ si } R_{i-1} > 0, \quad i = 2, \dots, m$$

où  $R_i$  désigne le nombre de trajectoires ayant atteint la frontière du niveau  $D_i$ .

### 3.4 Choix des paramètres importants

#### 3.4.1 les niveaux intermédiaires

Le choix des niveaux intermédiaires (et de leur nombre) est déterminant. En effet, pour que la simulation par sous-ensembles soit efficace, il ne faut pas que l'une des probabilités conditionnelles soit trop faible. En fiabilité des structures, l'évènement rare est le domaine de défaillance  $D = \{G(X) \leq 0\}$ , où  $G$  est la fonction d'état limite. En définissant  $D_i = \{G(X) \leq l_i\}$   $i = 1, \dots, m$  et  $l_1 > l_2 > \dots > l_m = 0$ , nous obtenons bien une suite d'ensembles emboîtés. Le choix de la suite décroissante  $(l_i)_{i=1, \dots, m}$  détermine les valeurs des probabilités conditionnelles : si la suite décroît trop lentement, les probabilités conditionnelles seront assez grandes et leur estimation ne demanderont pas un nombre de simulations  $n_i$  élevé, mais le nombre de niveaux intermédiaires  $m$  devra être élevé pour atteindre  $D$ , augmentant ainsi le nombre de simulations

total dans l'estimation de  $P_f$ . A l'inverse, si la suite décroît trop rapidement, le nombre de niveaux  $m$  sera plus faible mais le nombre de simulations pour estimer avec précision les probabilités conditionnelles sera sensiblement plus élevé. Le choix des niveaux intermédiaires est un compromis entre le nombre de simulations demandé à chaque étapes pour estimer avec précision les niveaux intermédiaires et le nombre de niveaux  $m$ .

La stratégie la plus simple est de choisir  $m$  et la suite  $(l_i)_{i=1,\dots,m}$  a priori mais il est alors difficile de contrôler les valeurs des probabilités conditionnelles, à moins d'avoir de solides informations sur le domaine de défaillance et sur la probabilité de rupture recherchée.

Une autre stratégie consiste à construire les niveaux intermédiaires de manière adaptative : la valeur des probabilités conditionnelles est fixée à l'avance  $p_i = p_0 \in ]0, 1[$ , un  $n$ -échantillon suivant  $f(\cdot; D_i)$  est généré et la valeur de  $l_i$  est fixée telle que  $\hat{p}_i = p_0$ . La figure 3.4.1 illustre ce principe. Le choix des niveaux intermédiaires est alors dépendant de l'échantillon généré (et ainsi de la probabilité de transition  $p^*$  choisie), ces niveaux variant donc à chaque nouvelle simulation.

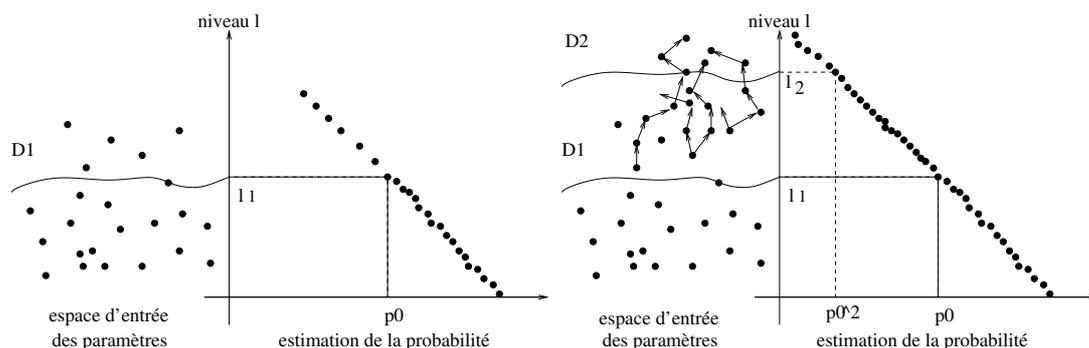


FIG. 3.4.1 – *Choix adaptatif des domaines intermédiaires*

### 3.4.2 La loi instrumentale

Le choix de la loi instrumentale joue un rôle crucial dans l'efficacité des méthodes MCMC, cette loi détermine l'état suivant de la chaîne de Markov. Idéalement, celle-ci doit couvrir le support de la loi stationnaire  $f(\dots, D_i)$  et doit aussi être une bonne approximation de  $f(\dots, D_i)$ .

Il est possible d'utiliser une loi symétrique centrée sur l'échantillon :  $p^*(\tilde{x}, x) = p^*(x, \tilde{x})$ , par exemple, une loi gaussienne ou uniforme centrée sur l'état courant. Ces lois devront couvrir suffisamment l'espace pour permettre le passage dans le domaine suivant, sans non plus revenir trop souvent dans le domaine précédent.

Un autre possibilité est d'utiliser une loi ne dépendant pas de l'échantillon :  $p^*(\tilde{x}, x) = h(\tilde{x})$ , cela nécessite une bonne connaissance de la loi cible  $f(\cdot, D_i)$  afin de choisir  $p^*$  aussi proche que possible de celle-ci.

### 3.5 Forme particulière du domaine de défaillance

Ici, nous nous plaçons dans le cas particulier où les domaines intermédiaires peuvent s'écrire sous la forme suivante (c'est le cas de l'application présentée au chapitre 4) :

$$D_i = \{(X_1, \dots, X_{d-1}, S) = (X, S), S - g(X) < l_i\}, \quad i = 1, \dots, m, \quad (3.5.1)$$

avec  $l_1 > l_2 > \dots > l_m = 0$  et  $g : x \in \Omega \subset \mathbb{R}^{d-1} \mapsto g(x) \in \mathbb{R}$ .

#### 3.5.1 Isolement d'une variable

Nous nous placerons dans le cas où  $X \in \mathbb{R}^{d-1}$  et  $S \in \mathbb{R}$  sont indépendants;  $X$  est distribué suivant la densité de probabilité multidimensionnelle  $f$  et  $S$  suivant la densité de probabilité unidimensionnelle  $w$ . Nous souhaitons exploiter au maximum l'information que nous avons sur la variable aléatoire  $S$ .

Les estimateurs des probabilités  $p_i = P(D_i|D_{i-1})$ ,  $i = 1, \dots, m$  de l'expression (3.1.1) peuvent alors s'écrire :

$$\hat{p}_i = \frac{1}{n} \sum_{k=1}^n \frac{W_i(x_k)}{W_{i-1}(x_k)}, \quad x_k \sim f(\cdot, D_{i-1}), \quad (3.5.2)$$

avec  $W_i(x) = \int \mathbf{1}_{D_i}(x, s)w(s)ds$  et  $f(x, D_{i-1}) = \frac{W_{i-1}(x)}{P(D_{i-1})}f(x)$ .  
 $f(\cdot, D_{i-1})$  s'exprime comme ci-dessus car nous avons

$$P((X, S) \in D_{i-1}, X \in A) = \int_A f(x, D_{i-1})P(D_{i-1})dx,$$

or,

$$W_{i-1}(x) = P((x, S) \in D_{i-1}) = P((X, S) \in D_{i-1}|X = x),$$

ainsi,

$$P((X, S) \in D_{i-1}, X \in A) = \int_A W_{i-1}(x)f(x)dx.$$

Pour simuler une chaîne de Markov de loi stationnaire  $f(\cdot, D_{i-1})$ , on utilise alors l'algorithme suivant :

Soit  $p^*$  choisie telle que :  $p^*(x, y) = p^*(y, x)$ .

Pour  $k = 0, 1, 2, \dots$  :

Étant donné un état  $x_k \in D_{i-1}$ ,

1. Générer un état « candidat »  $\tilde{x}_k$  :

Soit  $\xi_k \sim p^*(\cdot, x_k)$

Poser

$$\tilde{x}_k = \begin{cases} \xi_k & \text{avec la probabilité } \min \left\{ 1, \frac{f(\xi_k)}{f(x_k)} \right\} \\ x_{k-1} & \text{avec la probabilité } 1 - \min \left\{ 1, \frac{f(\xi_k)}{f(x_k)} \right\} \end{cases}$$

2. Choix de l'état suivant :

$$x_{k+1} = \begin{cases} \tilde{x}_k & \text{avec la probabilité } \min \left\{ 1, \frac{W_{i-1}(\tilde{x}_k)}{W_{i-1}(x_k)} \right\} \\ x_k & \text{avec la probabilité } 1 - \min \left\{ 1, \frac{W_{i-1}(\tilde{x}_k)}{W_{i-1}(x_k)} \right\} \end{cases}$$

Vérifions que cet algorithme crée une chaîne de Markov admettant pour loi stationnaire  $f(\cdot, D_{i-1})$ . Soit  $p(\cdot, \cdot)$  la densité de transition entre deux états d'une chaîne partant de  $D_i$ . Pour  $x_{k+1} \neq x_k$ , nous avons :

$$\begin{aligned} p(x_{k+1}, x_k) &= p^*(x_{k+1}, x_k) \min \left\{ 1, \frac{f(x_{k+1})}{f(x_k)} \right\} \min \left\{ 1, \frac{W_{i-1}(x_{k+1})}{W_{i-1}(x_k)} \right\} \\ &= \frac{f(x_{k+1})}{f(x_k)} \frac{W_{i-1}(x_{k+1})}{W_{i-1}(x_k)} p(x_k, x_{k+1}) \end{aligned} \quad (3.5.3)$$

Et on obtient, pour  $x_k \sim f(\cdot, D_{i-1})$  :

$$\begin{aligned} p(x_{k+1}) &= \int p(x_{k+1}, x_k) \frac{W_{i-1}(x_k)}{P(D_{i-1})} f(x_k) dx_k \\ &= \int p(x_k, x_{k+1}) \frac{W_{i-1}(x_{k+1})}{P(D_{i-1})} f(x_{k+1}) dx_k \\ &= \frac{W_{i-1}(x_{k+1})}{P(D_{i-1})} f(x_{k+1}). \end{aligned}$$

### 3.5.2 Propriétés de l'estimateur

Le premier estimateur  $\hat{p}_1$  est un estimateur classique d'une méthode de Monte Carlo direct : il est sans biais et converge presque sûrement vers  $P(D_1)$  quand  $n \rightarrow +\infty$ . Son coefficient de variation est donné par :  $\delta_1^2 = \frac{1 - P(D_1)}{P(D_1)n_1}$ .

Nous allons maintenant étudier les estimateurs  $\hat{p}_i$ ,  $i = 2, \dots, m$ , la chaîne de Markov nous permet d'obtenir des réalisations de variables aléatoires selon  $f(\cdot, D_{i-1})$  mais non indépendantes. La dépendance n'affecte pas la convergence presque sûre de  $\hat{p}_i$  vers  $P(D_i|D_{i-1})$  quand  $n_i \rightarrow +\infty$ .

Les estimateurs  $\hat{p}_i$ ,  $i = 2, \dots, m$  sont sans biais :

$$\begin{aligned} \mathbb{E}[\hat{p}_i] &= \int \frac{W_i(x)}{W_{i-1}(x)} \frac{W_{i-1}(x)}{P(D_{i-1})} f(x) dx \\ &= \frac{\int W_i(x) f(x) dx}{P(D_{i-1})} \\ &= P(D_i|D_{i-1}) \end{aligned}$$

Nous allons à présent étudier l'erreur quadratique moyenne de  $\hat{p}_i$  : soit  $n_i^c$  le nombre de chaînes de Markov simulées en parallèle, les états initiaux  $x_0^{(k)} \in D_{i-1}$ ,  $k = 1, \dots, n_i^c$  ne sont pas nécessairement tous différents. Chaque chaîne comporte  $n_i/n_i^c$  états.

On note  $I_{jk}^{(i)} = \frac{W_i(x_{jk}^{(i-1)})}{W_{i-1}(x_{jk}^{(i-1)})}$ , où  $x_{jk}^{(i-1)}$  désigne le  $k^{\text{ième}}$  état de la  $j^{\text{ième}}$  chaîne de Markov partant du domaine  $D_{i-1}$ . On suppose que  $\mathbb{E} \left[ \frac{W_i(x)}{W_{i-1}(x)} \frac{W_i(\tilde{x})}{W_{i-1}(\tilde{x})} \right] - P(D_i|D_{i-1})^2 = 0$  si  $x$  et  $\tilde{x}$  proviennent de chaînes différentes. En reprenant les résultats de l'article [AB 01], nous obtenons :

$$\begin{aligned} \mathbb{E} [(\hat{p}_i - p_i)^2] &= \mathbb{E} \left[ \left( \frac{1}{n_i} \sum_{j=1}^{n_i^c} \sum_{k=1}^{n_i/n_i^c} \frac{W_i(x_{jk}^{(i-1)})}{W_{i-1}(x_{jk}^{(i-1)})} - p_i \right)^2 \right] \\ &= \frac{R_i(0)}{n} (1 + \gamma_i), \end{aligned}$$

où :

- $\gamma_i = 2 \sum_{k=1}^{n_i/n_i^c-1} \left( 1 - \frac{kn_i^{(c)}}{n} \right) \frac{R_i(k)}{R_i(0)}$
- $R_i(k) = \mathbb{E} \left[ (I_{jl}^{(i)} - p_i)(I_{j,k+l}^{(i)} - p_i) \right] = \mathbb{E} \left[ I_{jl}^{(i)} I_{j,k+l}^{(i)} \right] - p_i^2$  est la covariance entre  $I_{jl}^{(i)}$  et  $I_{j,k+l}^{(i)}$  pour chaque  $l = 1, \dots, n_i/n_i^c$ .

### 3.6 Simulations sur des cas tests

Nous allons faire des essais sur des fonctions d'état limites analytiques.

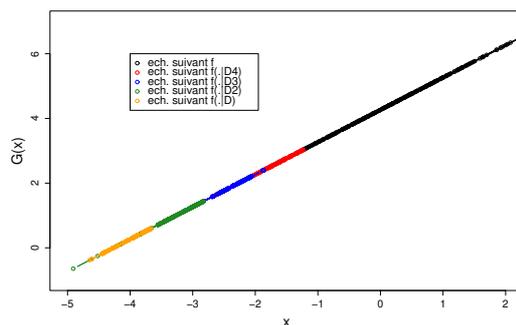
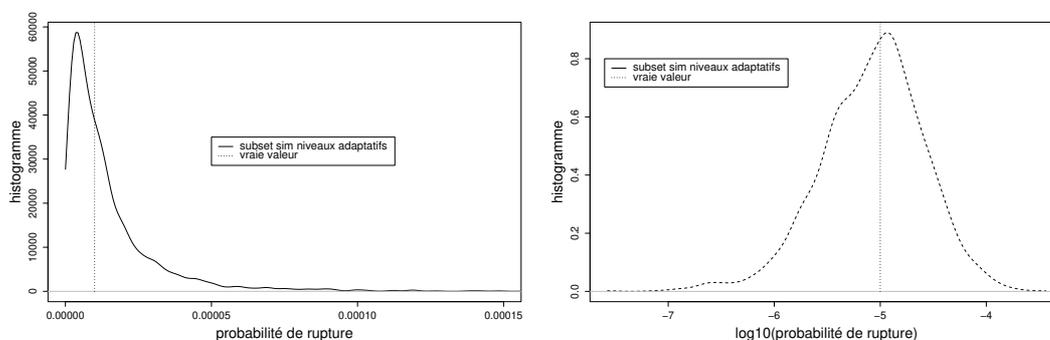
#### 3.6.1 Exemple 1

Nous reprenons l'exemple 1 en dimension 1 présenté en 2.4.1,  $X$  est une variable aléatoire de loi normale centrée réduite et la fonction d'état limite est définie par :

$$G(x) = x - \Phi^{-1}(10^{-5}) \tag{3.6.1}$$

Le graphique 3.6.1 illustre l'évolution des différentes chaînes de Markov de lois stationnaires  $f(\cdot, D_{i-1})$ ,  $i = 1, \dots, 5$ . La loi instrumentale est une loi symétrique : une gaussienne centrée sur l'état courant.

Les résultats présentés ci-dessus ont été obtenus à partir d'un nombre total d'états de 500. Les histogrammes des estimateurs sont tracés à partir de 2000 expériences.

FIG. 3.6.1 – *Évolution des chaînes de Markov, exemple 1d*FIG. 3.6.2 – *Histogramme de probabilités de rupture, exemple 1d*

### 3.6.2 Exemple 2

Sur cet exemple, la fonction d'état limite est maintenant définie par :

$$G(x) = -\sqrt{|x_1|} + 8 * \exp(x_2) \quad (3.6.2)$$

$X_1$  est une variable aléatoire de loi normale centrée réduite et  $X_2$  une variable aléatoire de moyenne 2 et de variance 1.

La figure 3.6.3 illustre l'évolution des échantillons de loi  $f(\cdot, D_{i-1})$ ,  $i = 1, \dots, 5$  vers le domaine de défaillance. La loi instrumentale est une loi symétrique : une gaussienne centrée sur l'état courant.

Les résultats présentés ci-dessus ont été obtenus à partir d'un nombre total d'états de 800. Les histogrammes des estimateurs sont tracés à partir de 2000 expériences.

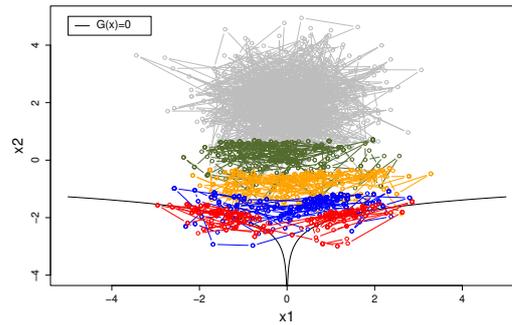


FIG. 3.6.3 – *Évolution des chaînes de Markov, exemple 2d*

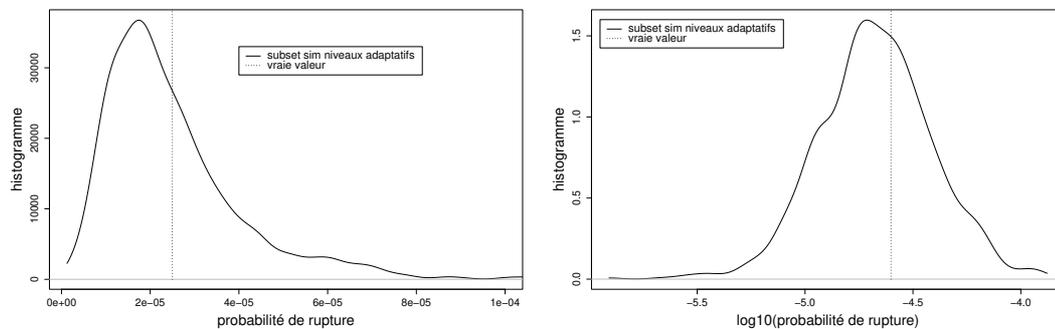


FIG. 3.6.4 – *Histogramme de probabilités de rupture, exemple 2d*

### 3.6.3 Exemple 3

On considère maintenant un exemple en dimension 4, la fonction d'état limite est donnée par :

$$G(x) = x_1^2 - x_2 + x_3x_4 + 10, \quad (3.6.3)$$

et  $X = (X_1, X_2, X_3, X_4)$  est un vecteur de lois normales centrées réduites indépendantes.

Les résultats présentés ci-dessus ont été obtenus à partir d'un nombre total d'états de 1000. Les histogrammes des estimateurs sont tracés à partir de 2000 expériences. La loi instrumentale est une loi symétrique : une gaussienne centrée sur l'état courant.

### 3.6.4 Exemple 4

Nous allons comparer sur cet exemple simple l'estimateur de la probabilité de défaillance obtenu avec les estimateurs des probabilités conditionnelles définis dans l'équation (3.5.2) et celui obtenu avec les estimateurs des probabilités conditionnelles définis dans l'équation (3.3.1).

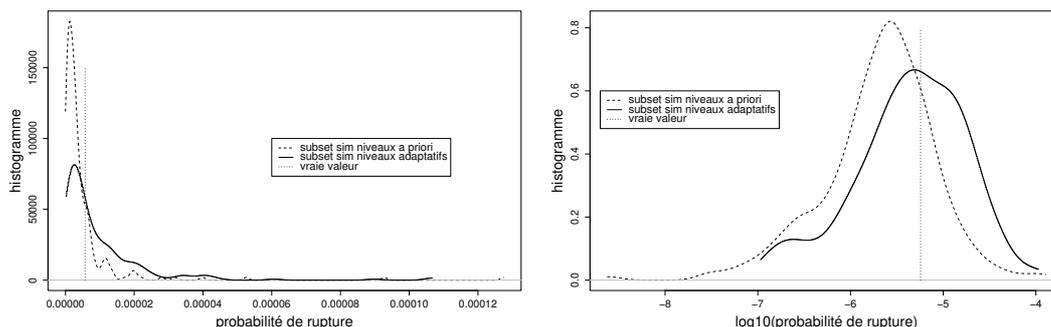


FIG. 3.6.5 – *Histogramme de probabilités de rupture, exemple 4d*

Le domaine de défaillance est défini par :

$$G(x, s) = 400 - 10x - s, \quad (3.6.4)$$

$X$  est une variable aléatoire de loi gaussienne de moyenne 80 et de variance 10 et  $S$  est une variable aléatoire de loi Weibull de module 5, de paramètre d'échelle 200 et de paramètre de localisation 0.

Les résultats présentés ci-dessus ont été obtenus à partir d'un nombre total d'états de 1000. Les histogrammes des estimateurs sont tracés à partir de 200 expériences. La loi instrumentale est une loi gaussienne centrée sur l'état courant.

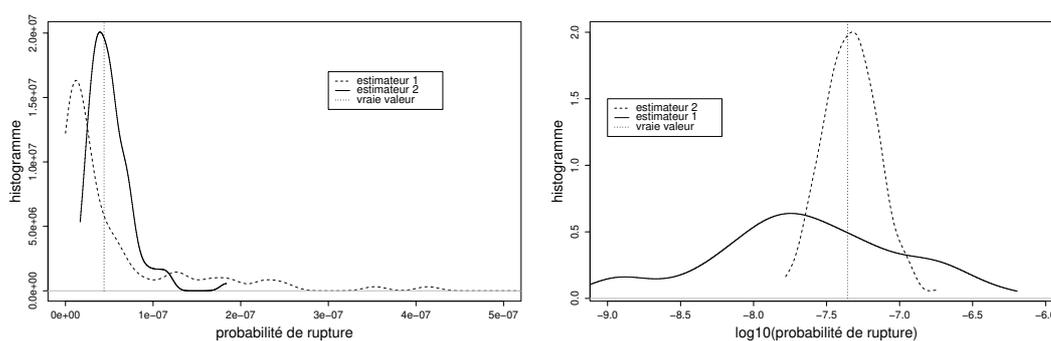


FIG. 3.6.6 – *Histogramme des estimateurs dans le cas où  $X$  et  $S$  sont indépendants*

Sur cet exemple, il apparaît clairement que l'estimateur 1 (3.5.2) donne de meilleurs résultats que l'estimateur 2 (3.3.1). Dans l'estimateur 1, c'est seulement  $x$  qui est exploré par la chaîne de Markov car on exploite toute l'information que l'on a sur  $s$ , et dans l'estimateur 2, c'est sur les variables  $x$  et  $s$  que ce fait l'exploration de la chaîne de Markov.

### 3.7 Conclusion

Le choix des niveaux a priori demande préalablement un réglage, ce qui augmente le nombre d'appels à la fonction d'état limite. On pourrait choisir un large nombre de niveaux pour estimer les probabilités conditionnelles avec précision, cela augmente aussi le nombre d'appels à la fonction d'état limite. Un choix adaptatif des niveaux est un bon compromis bien que ceci nécessite un choix judicieux de distribution instrumentale. En effet, en supposant que la loi instrumentale soit une loi symétrique (gaussienne ou uniforme centrée sur l'état courant), si l'écart-type pour la gaussienne ou « l'étendue » pour l'uniforme sont trop petits, le nombre de niveaux intermédiaires pour atteindre le domaine de défaillance sera très élevé, sous-estimant considérablement la probabilité de défaillance.

Dans le cas où le domaine de défaillance a une forme particulière (cf. § 3.5) et qu'une des variables est indépendante des autres, l'estimateur proposé au paragraphe 3.5.1 permet de l'isoler afin d'exploiter au maximum l'information que l'on a dessus.



## CHAPITRE 4

### APPLICATION AU COMBUSTIBLE HTR

#### 4.1 Contexte industriel

Cette section a pour objectif de présenter les réacteurs de type HTR et leur combustible. Nous citons les articles dont nous nous sommes inspirés pour présenter cette synthèse : [Bas ], [LP 02] , [LSE 02] et [Lan 02].

##### 4.1.1 Le réacteur à haute température

Depuis quelques années, les réacteurs à haute température (HTR) suscitent un regain d'intérêt dans plusieurs pays, mais c'est dans les années soixante que le développement des HTR a été amorcé. Une première vague de petits réacteurs a alors rapidement vu le jour : DRAGON au Royaume-Uni, en service en 1964, sans production d'électricité ; PEACH BOTTOM aux États-Unis, en service en 1967, de 40 MW de puissance électrique et l'AVR en Allemagne en 1967, de 15 MW de puissance électrique. Ils sont alors suivis de réacteurs de taille plus importante : FORT SAINT VRAIN aux États-Unis, en service en 1974 et THTR 300 en Allemagne, respectivement de puissance électrique 330 et 300 MW. A ce jour, ces réacteurs ont tous été arrêtés. Actuellement, deux réacteurs expérimentaux fonctionnent : HTTR au Japon, en service en 1998 et HTR 10 en Chine, en service en 2000, qui ne produisent pas d'électricité. Ces réacteurs expérimentaux servent surtout à consolider les acquis sur cette filière et permettre ainsi de maintenir un savoir-faire sur la technologie des HTR et de son combustible très spécifique.

Les HTR se caractérisent par :

1. un refroidissement par gaz inerte tel que l'hélium (on parle de réacteur à caloporteur gaz) ; parmi les gaz potentiellement utilisables, l'hélium présente en effet de grandes qualités d'échange et de transport de la chaleur et surtout la qualité d'être transparent aux neutrons ;
2. l'utilisation d'un combustible à particules particulièrement confinant et réfractaire (cf. § 4.1.2), il a été imaginé par les chercheurs d'Harwell dans les années cinquante.
3. l'emploi de grandes quantités de graphite comme modérateur, assurant une bonne stabilité mécanique et une grande inertie thermique tout en ayant de bonnes qualités neu-

troniques en fonctionnement nominal et accidentel. De plus, la densité de puissance volumique est faible (2 à 7 MW.m<sup>-3</sup> à comparer à 100 MW.m<sup>-3</sup> pour le parc actuel de REP).

La combinaison de ces trois spécificités conduit aux avantages suivants :

1. le fonctionnement à des températures élevées du caloporteur He (au-delà de 1000 °C pour le concept VHTR), ce qui permet des rendements thermodynamiques élevés ; cette caractéristique permet de nouvelles applications (production d'hydrogène en plus du courant électrique par exemple) ;
2. un niveau de sûreté intrinsèque unique dû à une faible densité de puissance, aux matériaux réfractaires et à une réactivité décroissante avec la température.

La figure 4.1.1 illustre le concept VHTR (Very High Temperature Reactor) d'AREVA.

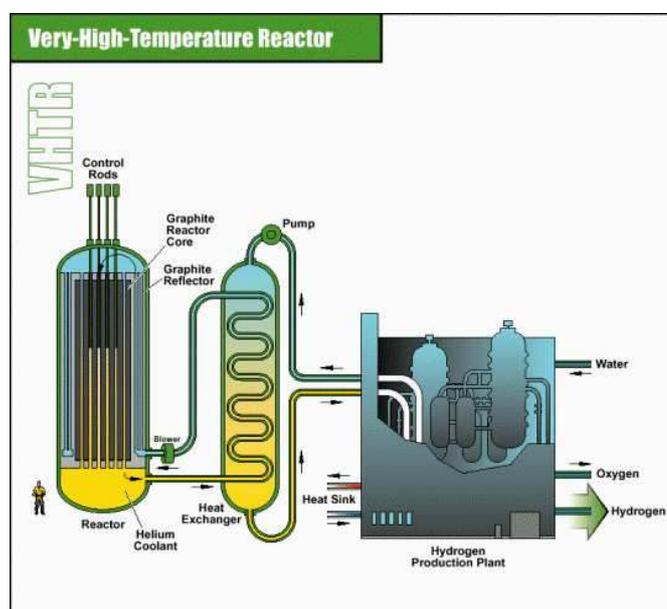


FIG. 4.1.1 – Réacteur à très haute température (concept VHTR d'AREVA)

A contrario, ce type de réacteur nécessite un volume de coeur important et est en conséquence peu compétitif avec les réacteurs de la filière REP, surtout pour des puissances installées élevées.

## 4.1.2 Le combustible à particules

### 4.1.2.1 Présentation du combustible

L'utilisation d'un combustible à particules a été l'innovation la plus décisive dans le concept des HTR. Ce combustible se présente sous la forme d'une petite sphère dont le diamètre est de l'ordre du millimètre (cf. figure 4.1.2 a). Cette sphère est composée d'un noyau de matière fissile enrobé de différentes couches de matériaux réfractaires. Dans les options aujourd'hui retenues, le noyau fissile est constitué d'oxyde d'uranium ( $\text{UO}_2$ ) et revêtu de 4 couches (particule TRISO pour TRistructural ISOtropic). Partant du noyau, se trouvent successivement (cf. figure 4.1.2 b) :

- une couche de carbone pyrolytique très poreux ( $> 50\%$  de porosité), le buffer, servant de réservoir pour les gaz de fission relâchés par le noyau et de protection des couches suivantes; il permet également d'accommoder le gonflement sous irradiation du noyau.
- une couche de carbone pyrolytique « dense » ( $\sim 15\%$  de porosité), l'IPyC, contribuant à l'étanchéité de la particule vis-à-vis des gaz de fission et à la tenue mécanique de la particule, elle facilite aussi le dépôt de la couche suivante;
- une couche en carbure de silicium, le SiC qui est la principale structure résistante de la particule; elle sert à assurer l'étanchéité de la particule vis-à-vis d'autres produits de fission tels que les PF volatiles et métalliques;
- une couche de carbone pyrolytique dense, l'OPyC, protégeant la couche SiC; elle permet de retenir aussi les produits de fission des particules à couches SiC non étanche.

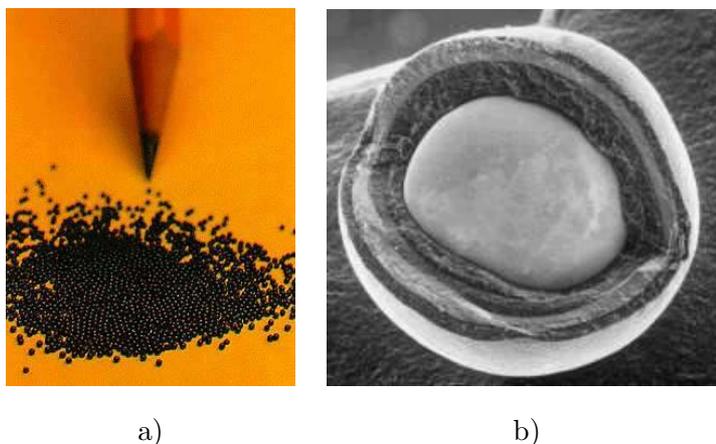


FIG. 4.1.2 – Le combustible à particules est un combustible très fractionné dont les propriétés s'évaluent de manière statistique a). Chaque particule (de  $900\mu\text{m}$  de diamètre) est constituée d'un noyau combustible enrobé de différentes couches réfractaires b).

Cette conception possède de grands intérêts :

- un cœur de réacteur va contenir des quantités considérables de particules (de  $10^9$  à  $10^{11}$ ) ; c'est statistiquement que vont s'apprécier ses performances ;
- elle permet un excellent confinement des produits de fission (l'étanchéité est assurée en cas d'accident de refroidissement (dépressurisation primaire) jusqu'à  $1600\text{ °C}$ ), et permet ainsi d'accéder à des températures en fonctionnement nominal élevées ;
- elle offre une très grande variété de combinaisons des matériaux et aussi des paramètres géométriques ou physiques ; le noyau peut être de l'uranium naturel ou enrichi, du thorium ou plutonium, sous forme d'oxyde ou de carbure ; la couche SiC peut être remplacée par du carbure de zirconium, plus réfractaire que le SiC ;
- les qualités de l'enrobage permettent d'envisager un stockage en limitant les opérations de conditionnement supplémentaires.

Les particules ne sont pas disposées librement dans le cœur du réacteur mais agglomérées dans une matrice en graphite pour former des corps cylindriques ou sphériques aisément manipulables :

- les compacts sont de petits cylindres de 1 à 2 centimètres de diamètre et de 5 à 6 centimètres de long ; ils sont insérés à l'intérieur de gros blocs prismatiques en graphite à section hexagonale (30 centimètres environ de diamètre pour 80 centimètres de hauteur) ; ces blocs sont percés de canaux pour le passage du gaz caloporteur ; (cf. figure 4.1.3 a) et b))
- les boulets (cf. figure 4.1.3 b)) sont des sphères en graphite d'environ 6 centimètres de diamètre ; le cœur est alors constitué d'un empilement « en vrac » de ces boulets, le caloporteur circulant dans l'espace laissé libre par cet empilement.

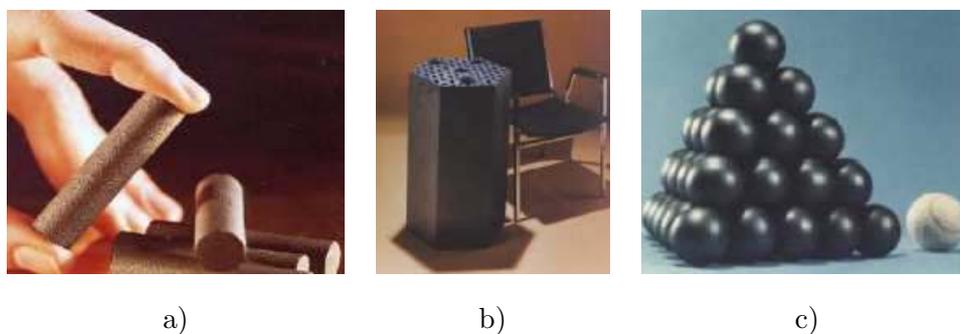


FIG. 4.1.3 – Les assemblages combustibles se présentent sous la forme de compacts a) insérés dans des blocs b) ou sous la forme de boulets c).

#### 4.1.2.2 Comportement sous irradiation

Les principaux phénomènes qui régissent le comportement de la particule sous irradiation sont :

1. le comportement des couches PyC sous irradiation ; les couches IPyC et OPyC ont tendance à se densifier d'abord identiquement dans toutes les directions pour une faible fluence neutronique ; puis elles tendent à gonfler radialement et se densifient orthoradialement pour des fluences plus élevées, cette déformation est accommodée par la création de contraintes importantes et généralement par le fluage sous irradiation du PyC. Enfin, le coefficient d'anisotropie augmente de par les contraintes générées.
2. la mise en pression interne de la particule par la production de gaz de fission et d'oxyde de carbone dans le noyau et leur relâchement dans le buffer ; la pression des gaz augmente avec le taux de combustion et met progressivement la couche SiC en tension (cf. 4.1.4) ;
3. la diffusion de certains produits de fission radioactifs ( $^{85}\text{Kr}$ ,  $^{137}\text{Cs}$ ,  $^{90}\text{Sr}$ , ...) à travers les différentes couches. Le SiC est en général une barrière efficace pour la plupart d'entre eux .

L'objectif du combustible à particules est de supporter des hautes températures et des taux de combustion élevés (15% à 20% FIMA) tout en garantissant un taux de particules rompues très faible et ainsi un taux de relâchement de produits de fission très limité, en situation nominales et accidentelles. Le comportement sous irradiation de la particule peut être induit schématiquement les phases suivantes :

1. à taux de combustion faible (< 2% FIMA), le buffer subit une forte densification ce qui entraîne une augmentation du volume libre dans la particule par la création d'un jeu autour du buffer, les couches PyC se densifient et comme elles sont soudées à la couche SiC, elles sont mises en tension et la couche SiC est mise en compression (cf. figure 4.1.4 b)) ; il peut éventuellement y avoir rupture du PyC ;
2. à taux de combustion modéré (entre 4% et 8% FIMA), le gonflement du noyau entraîne une diminution du volume libre et du jeu, les gaz de fission sont produits par le noyau et la pression interne de la particule augmente. Les contraintes de tension dans les couches PyC et donc de compression de la couche SiC s'atténuent (cf. figure 4.1.4 c)) ;
3. à partir d'un taux de combustion plus élevé (> 10% FIMA), la pression interne appuyant sur les couches, la contrainte orthoradiale dans la couche SiC peut s'inverser. Cette contrainte va continuer à augmenter avec le taux de combustion et peut éventuellement provoquer la rupture de cette couche (le SiC ne peut rompre qu'en tension), ce qui

entraînerait la rupture des couches PyC et un relâchement très important des produits de fission hors de la particule(cf. figure 4.1.4 d)).

La figure 4.1.5 illustre l'évolution temporelle des contraintes des couches différentes couches principales : IPyC, SiC et OPyC.

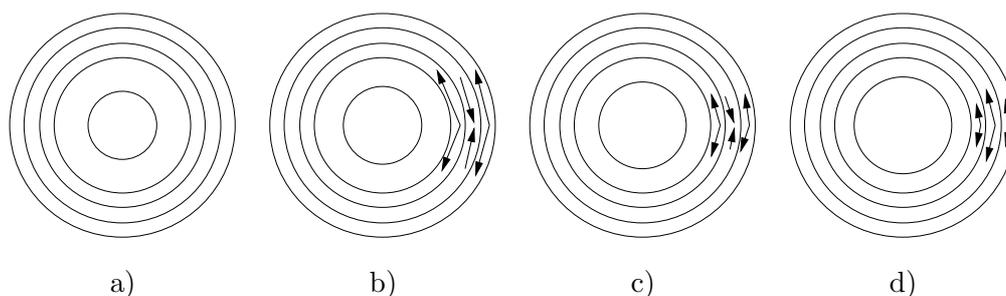


FIG. 4.1.4 – Comportement sous irradiation d'une particule à différents taux de combustion [LP 02] ; a)  $t = 0$ , b)  $\sim 1\%$  FIMA, c)  $\sim 5\%$  FIMA, d)  $\sim 12\%$  FIMA. % FIMA : pourcentage d'atomes lourds fissionnés par rapport aux atomes lourds initiaux.

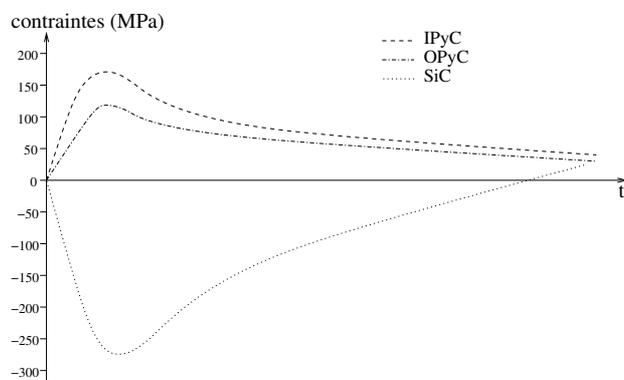


FIG. 4.1.5 – Évolution des contraintes orthoradiales dans les couches IPyC, SiC et OPyC au cours du temps.

Le combustible à particules a été éprouvé par de nombreuses irradiations dans des réacteurs expérimentaux et des essais ont été réalisés pour simuler les conditions accidentelles. Dans le cas d'un accident de dépressurisation d'un réacteur HTR, la température du cœur peut atteindre  $1600^{\circ}\text{C}$ , la particule doit être capable de retenir les produits de fission jusqu'à cette température. Les retours d'expériences montrent que la couche SiC joue en général parfaitement son rôle jusqu'à  $1600^{\circ}\text{C}$  ; au-delà, la corrosion du SiC par certains PF favorise le relâchement

des produits de fission et à partir de 2200°C, la couche SiC se décompose complètement, entraînant un relâchement quasi total des produits de fission créés.

#### 4.1.2.3 Les modes de ruptures de la particule

Les principaux modes de rupture des particules sont aujourd'hui bien établis. Parmi ceux pouvant augmenter considérablement le taux de produits de fission relâché (rupture de la couche SiC), nous retrouvons :

1. l'éclatement de la particule due à une surpression interne générée par les gaz de fission et oxydes de carbone relâchés par le noyau (cf. figure 4.1.6 a) ) ;
2. la rupture des couches de PyC ; les couches PyC peuvent être fissurées initialement (défaut de fabrication par exemple) et ne remplissent plus pleinement leur rôle (cf. figure 4.1.6 b) ) ; la rupture précoce de l'IPyC due à une vitesse de densification trop rapide peut entraîner quasi systématiquement la rupture du SiC ; la couche OPyC peut rompre à cause d'une trop forte interaction mécanique avec la matrice graphite du compact ou du boulet et fragilise le SiC (cf. figure 4.1.6 c) ) ;
3. l'effet amibe, c'est la migration du noyau sous l'effet d'un trop fort gradient thermique, il peut provoquer une destruction complète de la particule si le noyau atteint le SiC (cf. figure 4.1.6 d) ) ; l'effet amibe peut être évité grâce à un élément combustible limitant les gradients thermiques ;
4. la corrosion de la couche SiC ; certains produits de fission (palladium, lanthanides et le CO) peuvent corroder et détériorer le SiC ; C'est un mécanisme de rupture de fin de vie (cf. figure 4.1.6 e) ) .

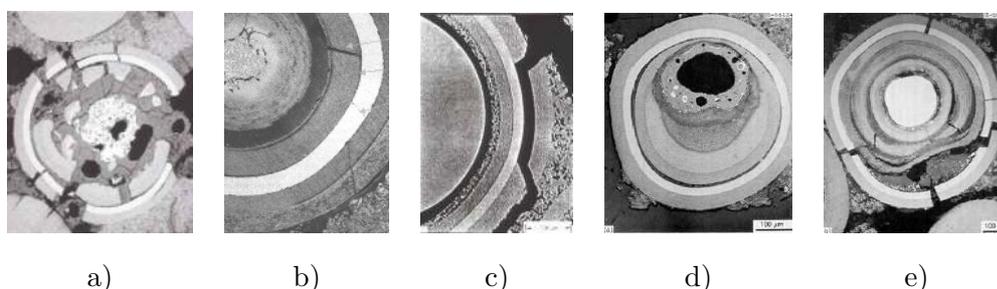


FIG. 4.1.6 – a) Rupture mécanique par pressurisation, b) rupture du PyC par interaction entre la matrice et la particule, c) fissuration du PyC et du SiC, d) rupture par effet amibe, e) rupture du SiC par corrosion par les produits de fission.

## 4.2 Codes modélisant le comportement d'une particule

Dans cette partie, nous allons tout d'abord décrire brièvement le code de calcul ATLAS, à savoir une simulation du comportement sous irradiation d'une particule, puis nous ferons un point particulier sur les méthodes statistiques utilisées pour calculer la probabilité de défaillance des particules. Ensuite, nous passerons sommairement en revue les différentes approches des autres codes utilisés dans le monde.

### 4.2.1 Le code ATLAS

ATLAS est une application dédiée à la simulation du comportement du combustible des réacteurs HTR lors de l'irradiation. Pour une présentation détaillée du code, nous pouvons nous référer à [Mic 03, Lai 05]. ATLAS est réalisé au CEA avec le soutien d'AREVA, qui utilise le code. La modélisation est développée dans le cadre d'un travail coopératif européen, le projet RAPAHÉL. ATLAS est un outil pour aider au dimensionnement du combustible. Il doit être validé par une confrontation des résultats des calculs aux résultats expérimentaux. Il pourra ensuite être aussi utilisé pour le licensing (constitution du dossier permettant d'obtenir l'autorisation, délivrée par l'autorité de sûreté du pays, de démarrer un réacteur de ce type). Il est déjà impliqué dans des comparaisons de résultats avec les codes concurrents présentés au paragraphe 4.2.2.1, notamment dans le cadre de l'AIEA.

Comme les particules sont de très petite taille (cf. 4.1.2), presque aucun résultat expérimental représentatif n'est obtenu à leur échelle. Ainsi, la validation doit être effectuée en se basant sur les résultats obtenus sur une population de particules, principalement la diffusion et le taux de rupture, les deux étant liés.

#### 4.2.1.1 Modélisation de comportement d'une particule

ATLAS inclut la connaissance que l'on a des phénomènes se déroulant au niveau du combustible, traduits sous forme de modèles numériques. Ce code comporte de nombreux modèles. La thermique, la mécanique et la diffusion des produits de fission sont résolus par éléments finis. La thermique et la mécanique sont isotropes et quasi stationnaires, avec un fluage dû à l'irradiation. La diffusion des PF est en début de transitoire, puisque les PF n'ont pas le temps de sortir de la particule. A cela s'ajoute divers modèles empiriques :

1. des modèles traduisant les déformations imposées par l'irradiation : les gonflements du noyaux dus à l'augmentation du nombre d'atome après fission de l'uranium, ainsi que aux gas de fission générant des bulles et la densification des matériaux de par la migration des bulles. Ces déformations sont orthotropes.

2. le relâchement des gaz de fission dû à une diffusion, puis une percolation des gaz de fission à travers les grains et joints de grain ;
3. un transfert de matière de la partie chaude du buffer vers la partie froide, générant un déplacement du noyau (effet amibe) ;
4. une corrosion de la couche de SiC, se traduisant par une perte locale des propriétés mécaniques de la couche ;
5. un calcul de la chaleur et de la quantité des produits de fission d'intérêt produits par la fission ;
6. l'équilibre chimique du CO et du CO<sub>2</sub> ;
7. des modifications des propriétés des couches lors de l'irradiation.

Comme ATLAS utilise les éléments finis pour les phénomènes où intervient la géométrie, il permet d'effectuer la résolution sur plusieurs géométries et types d'éléments. Ainsi, ATLAS réalise des calculs en 1D (voir figure 4.2.1), pour les particules sphériques, en 2D (voir figure 4.2.2) pour étudier certains défauts, et en 3D pour l'étude de l'interaction entre la particule et la matrice (voir figure 4.2.2). Lors d'une simulation 1D, la particule est supposée parfaitement sphérique et l'effet de la matrice n'est pas pris en compte : pas d'effet mécanique, pas de résistance à la diffusion. Les événements étudiés en 2D sont des défauts de sphéricité, des ruptures d'une couche, un gradient de sollicitation. Comme les temps de calcul sont fortement croissants avec le nombre de mailles de la représentation de l'objet, seuls les résultats 1D sont utilisables facilement lors d'une étude statistique et le nombre de simulation est déjà bien limité (quelques centaines de simulations en une nuit).

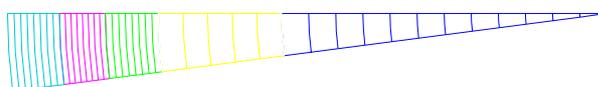


FIG. 4.2.1 – Géométrie 1D du code ATLAS

#### 4.2.1.2 Historique d'irradiation

L'historique d'irradiation est une entrée du code ATLAS, et permet de caractériser le réacteur et l'expérience. Cet ensemble de paramètres décrit les sollicitations extérieures auxquelles est soumise la particule. L'historique d'irradiation prend en compte de façon simplifiée tout le cycle de vie des particules, de la fabrication jusqu'à son taux de combustion final. Il est constitué par les évolutions temporelles de quatre quantités :

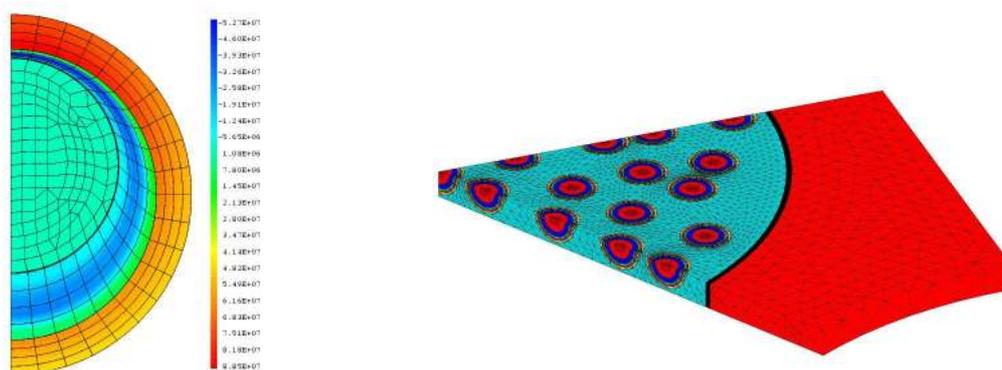


FIG. 4.2.2 – Géométrie 2D et 3D du code ATLAS - à gauche : simulation de l'effet amibe (contraintes), à droite : compact 3D.

- la température sur la face externe de la particule ;
- la pression exercée par le fluide caloporteur sur la face externe de la particule ;
- le flux rapide ;
- la vitesse de fission.

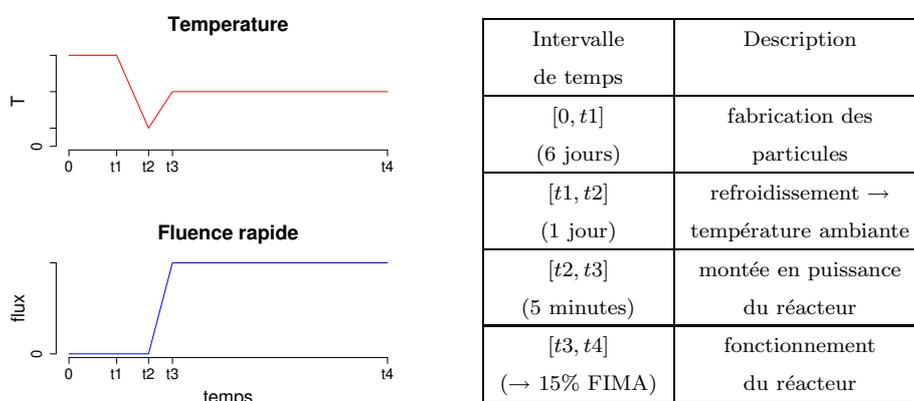


FIG. 4.2.3 – Exemple d'historique d'irradiation - Les courbes sont les évolutions temporelles de la température d'irradiation et de la fluence rapide. Les intervalles de temps de l'historique sont renseignés dans le tableau.

#### 4.2.1.3 Calcul de la probabilité de rupture

Nous avons vu au paragraphe 4.1.2 que l'objectif du combustible à particules est de garantir un taux de relâchement des produits de fission hors de la particule très limité, en situation nominales et accidentelles. Or, la principale barrière de ces produits de fission est la couche SiC. Nous nous intéresserons donc principalement à la rupture de cette couche céramique.

Nous supposons que la loi de rupture des céramiques, le SiC en particulier, suit une loi de Weibull. Cette loi est très utilisée pour caractériser la fiabilité des matériaux fragiles. La rupture est entièrement liée à la contrainte nécessaire pour propager une fissure ou un défaut, c'est la contrainte de rupture. La probabilité de défaillance du SiC en fonction de la contrainte orthoradiale dans cette couche sera alors donnée par la fonction de répartition de la loi de Weibull :

$$F(\sigma) = 1 - \exp \left\{ - \frac{(\sigma - \sigma_u)^m}{\sigma_o} \right\}, \quad (4.2.1)$$

où  $\sigma_o$  est le paramètre d'échelle,  $m$  le module de Weibull ( $m > 1$  correspond à un matériau qui se dégrade avec le temps;  $m < 1$  à un matériau qui se bonifie avec le temps et  $m = 1$  à un matériau sans usure) et  $\sigma_u$  paramètre de localisation ou de repérage. Dans notre cas, c'est le seuil en dessous duquel il n'y a pas de rupture. Nous prendrons  $\sigma_u = 0$  car nous supposons que la rupture peut avoir lieu dès que le SiC est en tension.

Le code ATLAS fournit à chaque instant de l'irradiation les contraintes orthoradiales dans les couches IPyC, SiC et OPyC. La probabilité de rupture moyenne est alors :

$$P_f = \int F(g(x))f(x)dx, \quad (4.2.2)$$

où  $g$  représente la valeur de la contrainte orthoradiale dans le SiC calculée par le code,  $x$  les paramètres d'entrée du code (cf. § 4.3.1) et  $f$  leur densité de probabilité.

**Remarque :**

En fait nous avons isolé la variable contrainte à rupture (cf. chapitre 1 § 1.3.2), celle-ci est indépendante des variables d'entrées du code. Nous sommes passés de l'intégrale

$$P_f = \int \mathbf{1}_{\{\sigma_R - g(x) \leq 0\}}(\sigma_R, x) f(x) w(\sigma_R) dx d\sigma_R \quad (4.2.3)$$

à l'intégrale (4.2.2),  $w$  la densité de la loi de weibull et  $\sigma_R$  la contrainte à rupture et  $\{\sigma_R - g(x) \leq 0\}$  est le domaine de défaillance.

Les méthodes du tirage d'importance adaptatif paramétrique décrites au paragraphe 2.3 ont été intégrées à ATLAS et peuvent être utilisées pour le calcul de la probabilité de défaillance (4.2.2), c'est en fait une combinaison d'une méthode de Monte Carlo conditionnel et du tirage d'importance. Il est aussi possible d'utiliser la méthode de simulation multi-niveaux utilisant les algorithmes de Monte Carlo par Chaîne de Markov (ch. chapitre 3) pour le calcul de la probabilité de défaillance et même celui introduit au paragraphe 3.5 car le domaine de défaillance a l'expression requise.

### 4.2.2 Les codes internationaux

Beaucoup de pays s'intéressent aux réacteurs HTR. Plutôt que de développer leur propre code, certains pays utilisent les codes anciens tels que PANAMA ou STRESS3 ou des variantes, qu'ils ont pu obtenir à partir des nombreux accords commerciaux. Il y a beaucoup d'autres échanges dans ce domaine. Depuis les années 90, le développement de la modélisation des particules HTR au niveau Européen s'est mutualisé. Dans le tableau 4.2.1, nous pouvons voir quelques uns des codes internationaux.

La conception et le licensing des réacteurs se basent très souvent sur des prédictions de comportement obtenues avec des codes. Les ruptures étant rares, il est très difficile et coûteux en expériences d'observer les particules indépendamment pour en déduire les causes de défaillance. Le code de simulation lui-même représente par contre des dépenses faibles. Les colloques et réunions, par exemples celles organisées par l'AIEA pour des mises en commun de données, sont des occasions de présenter les codes pour s'affirmer comme un acteur important.

pays	organisme	code
USA	INEEL	PARFUME
USA	MIT	TimCoat
Allemagne	FZJ	PANAMA
Angleterre	BNFL	STRESS3
Japon	JAERI	sans nom
Russie	VNIINM	GOLT
France	CEA	ATLAS

TAB. 4.2.1 – *Quelques uns des codes internationaux modélisant le comportement sous irradiation d'une particule*

Il est observé expérimentalement que les particules ne cassent pas toutes en même temps, ce qui amène à une modélisation probabiliste du phénomène, à la fois à travers une loi probabiliste de rupture, mais aussi en s'intéressant à la variabilité des conditions. Les codes cherchant à représenter la réalité, il leur est donc demandé d'arriver à représenter et à terme prédire le nombre de ruptures observées expérimentalement, indépendamment du contenu de la modélisation. Comme le nombre de ruptures dans des conditions expérimentales bien documentées est faible, la marge de variation des modèles est grande. Il est ainsi possible d'observer des modélisations assez différentes en ce qui concerne la physique et surtout la statistique de la rupture des particules. Ceci est aussi dû à la grande fiabilité des particules sur des petites ex-

périences, la difficulté à compter le nombre de particule défaillante sur les expériences à grande échelle, les difficultés à caractériser les particules et la forte variabilité de leurs performances si la fabrication a quelque peu changé.

#### 4.2.2.1 Modélisations basées sur une seule particule

Dans le passé, les allemands ont développé complètement un réacteurs prototype HTR, c'est à dire qu'ils ont construit et fait accepter le démarrage du réacteur. Pour arriver à ces fins, ils ont développé le code PANAMA [VN 90] qui repose sur le calcul de la contrainte dans la couche SiC due à la pression des gaz de fission, en approximation couche fine. La densification du pyrocarbone n'est pas prise en compte, ce qui constitue une différence majeure avec les autres codes présentés ici. Avec cette hypothèse, la pression SiC est toujours positive. En associant une probabilité de rupture à cette contrainte, on obtient une probabilité toujours croissante. La probabilité de rupture est celle de la particule moyenne. Un exemple est donné figure 4.2.4. Cette modélisation est certes très simple, mais comme elle a été validée et utilisée pour un prototype licencié, elle a une grande longévité.

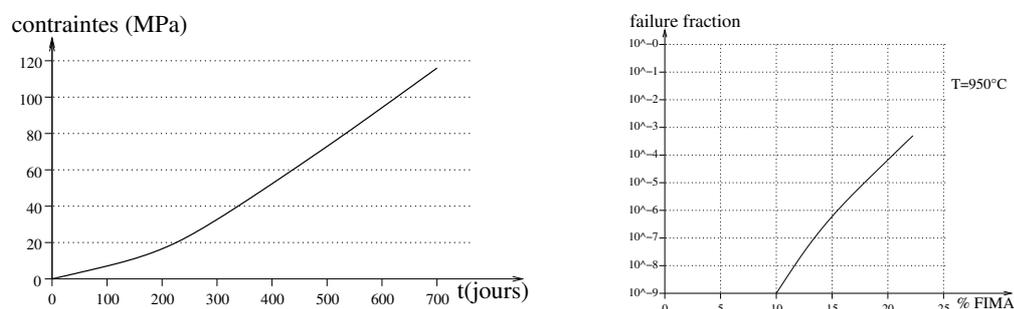


FIG. 4.2.4 – Évolution de la contrainte dans le SiC (à gauche) et de la probabilité de rupture (à droite) en fonction du temps par une formulation type PANAMA.

La plupart des codes plus récents prennent en compte plus précisément les principaux phénomènes, dont la densification sous irradiation du pyrocarbone. Cette densification crée des contraintes importantes sur le SiC, qui sont relaxées par le fluage (cf § 4.1.2.2). Ils sont généralement basés sur une résolution analytique de la mécanique (de l'élasticité) dans la géométrie sphérique des particules. Le fluage étant dépendant du temps, celui-ci est traité sous forme d'incrément [Wal 72] ou bien en écrivant tous les paramètres évolutifs sous forme de série entière [MB 93]. Les équations sont souvent utilisées de façon approchée en supposant le SiC rigide [Mar 73].

JAERI, par exemple, utilise un modèle analytique simple [SS 96]. La probabilité de rupture

est obtenue en utilisant l'évolution de la contrainte de la particule moyenne avec une contrainte à rupture dans le SiC aléatoire et de moyenne décroissante. L'évolution de la probabilité de rupture d'une particule est estimée en fonction de la température et de l'épaisseur du buffer. Aucune méthode statistique n'est utilisée (les paramètres de la particules sont fixés). Notons qu'il est supposé que le PyC ne rompt pas avant que le SiC ne soit lui-même rompu. Sur la base du design du HTTR, les résultats de JAERI sont comparés avec ceux de Verfondern (PANAMA) [Sa 01]. Dans des conditions normales, il est bien observé que le SiC ne se retrouve pas en tension pour une particule intacte et cela ne peut donc conduire à l'évolution observée des ruptures. Il est donc admis que certaines particules démarrent avec un SiC cassé, la couche OPyC retient les produits de fission jusqu'à sa rupture. Celle-ci est en tension et peut casser à tout moment au cours de l'irradiation.

Pour le code GOLT [Ga 02], la probabilité de rupture d'une particule est calculée de façon analytique mais le code ne travaille pas sur une population. Les paramètres dont l'influence est présentée sont le diamètre du noyau, l'épaisseur du Buffer et des couches IPyC et SiC. Comme les probabilités de ruptures ne sont pas encore réalistes, l'analyse statistique est présentée comme l'évolution prochaine du code.

Avec les lois prises communément, le pyrocarbone obtient son maximum de contrainte assez tôt dans l'irradiation et le SiC reste sous contrainte négative pendant longtemps. Ces modèles ont du mal à retrouver le comportement en irradiation où les ruptures sont très réparties.

#### 4.2.2.2 Modèles analytiques utilisant des méthodes statistiques

Martin [Mar 79] est dans les premiers à considérer des dispersions sur les données d'entrée. Il insiste clairement sur l'importance de la variabilité des données des particules. Il développe le code STRESS3 à partir de la formulation mathématique de Walther. Il propose ensuite :

1. d'utiliser une méthode de Monte Carlo directe sur un modèle simpliste, qui vient de Williamson et Horner [WH 71] et qui repose sur la formule :

$$P = \frac{2}{r} \left( t_5 \sigma_5 - \frac{t_4 \dot{g}_4}{K_4(1 - \nu_4)} - \frac{t_6 \dot{g}_6}{K_6(1 - \nu_6)} \right) \quad (4.2.4)$$

où – 4,5 et 6 correspondent à IPyC, SiC et OPyC respectivement ;

- $\sigma$  : contraintes ;
- $K$  : coefficient de fluage ;
- $\nu$  : coefficient de poisson pour le fluage ;
- $t$  l'épaisseur ;
- $P$  la pression, elle-même donnée par un modèle simple.

La contrainte à rupture est tirée aléatoirement suivant une loi de Weibull et le moment de la rupture en est déduit.

2. de remplacer le code par son développement à l'ordre 1 autour de la particule moyenne. Le code étant alors linéaire, et les paramètres d'entrée gaussien, la réponse est gaussienne. On utilise alors l'équation de propagation d'erreur autour de la particule moyenne. Une relation linéaire est aussi prise en compte entre le burn-up et la contrainte. La densité de probabilité de rupture à un burn-up  $b$  s'écrit (en adimensionné) :

$$p(b) = \int_0^\infty w(\sigma)g(\sigma - b)d\sigma \quad (4.2.5)$$

Le taux de rupture s'écrit alors :

$$P_f(b) = \int_{-1}^b p(x)dx \quad (4.2.6)$$

Dans [Mar 02], Martin précise que pour l'expérience HRB22 ou pour l'expérience 91F-1A, les ruptures ne peuvent pas être expliquées par le comportement de la particule moyenne mais uniquement en considérant la dispersion des caractéristiques.

Bongartz [Bon 80] au KFA à Julich utilise aussi le modèle de Walther, dans l'approximation du SiC rigide. Il présente 2 modèles statistiques :

1. Le modèle direct : 3 paramètres sont extraits dont l'épaisseur du buffer qui est le plus important. Pour une contrainte donnée  $\sigma_i$ , et avec les deux autres paramètres fixés, l'épaisseur du buffer qui y correspond  $Z_i(x, y)$  est déterminée par interpolation linéaire . La probabilité d'avoir une contrainte supérieure à la valeur considéré s'écrit :

$$G_i(\sigma_i) = \int_0^\infty H_x \int_0^\infty H_y \int_0^{Z_i(x,y)} H_z dz dy dx \quad (4.2.7)$$

avec  $H_x$  la densité de répartition du paramètre  $x$ . Avec plusieurs contraintes, on extrait le polynôme  $G(\sigma)$  qui passe par les différents points obtenus. Finalement, la probabilité moyenne de rupture s'écrit :

$$P_f = \int_0^\infty w(\sigma)G(\sigma)d\sigma \quad (4.2.8)$$

avec  $w$  la densité de la loi de Weibull correspondant à la contrainte à rupture du SiC.

2. Propagation d'erreur : la distribution de la contrainte est considérée comme gaussienne et la probabilité moyenne de rupture est directement déduite. La moyenne de la gaussienne des contraintes est supposée être la contrainte de la particule moyenne et l'écart type  $\mu$  est supposé être :

$$\mu = \sqrt{\sum_1^N \left(\frac{\partial \sigma}{\partial x_i} \mu_i\right)^2} \quad (4.2.9)$$

ou  $\mu_i$  est l'écart-type de la distribution du paramètre  $i$ .  $N$  est le nombre de paramètres. La probabilité moyenne de rupture  $P_f$  s'écrit :

$$P_f = \int_0^{\infty} w(\sigma)\varphi(\sigma)d\sigma \quad (4.2.10)$$

où  $w$  est la densité de probabilité de la contrainte à rupture (suivant une loi de weibull et  $\Phi$  est la fonction de répartition de la gaussienne considérée.

Avec l'amélioration des temps de calculs due aux progrès des ordinateurs, ce sont des méthodes de Monte Carlo sur les modèles analytiques qui vont être en général utilisés par la suite, généralement un Monte Carlo direct. Par exemple, TIMCOAT [WBM 04] utilise une formulation analytique à partir d'un développement en série infinie des termes en fonction du temps puis une méthode de Monte Carlo direct. Pour chaque particule, les caractéristiques, dont sa contrainte à rupture, puis ses conditions d'historique pour chacun de ses passages dans le réacteur (type PBMR) sont déterminées aléatoirement. L'effet de la rupture de l'IPyC sur la probabilité de rupture du SiC est prise en compte analytiquement à partir d'un modèle de rupture et de la variable  $K_1C$ . Lors d'un calcul sur des conditions de réacteur type PBMR [WBD 04], avec 1000000 de simulations, le résultat garde un écart type sur l'erreur de  $5.10^{-5}$  pour la probabilité moyenne de  $3.4.10^{-5}$ . Le code GOLT utilise pour le moment une méthode de Monte Carlo conditionnel, mais l'intégration du tirage d'importance est envisagé.

#### 4.2.2.3 Utilisation des éléments finis dans la modélisation

Bien que les formulations analytiques soient assez précises. Les simulations par éléments finis restent la solution de référence pour traiter les défauts localisés ou l'effet des défauts de symétrie sphérique. Dans [Ben 91], il est d'ailleurs montré que la modélisation par EF est plus précise, même dans le cas sphérique. Dans le cas de simulation de défauts, les contraintes obtenues sont fortement inhomogènes. La probabilité de rupture peut être calculée à partir des contraintes obtenues par EF en considérant la rupture comme un processus de Poisson. C'est à dire que l'on fait sur le maillage une intégration volumique du terme  $(\sigma/\sigma_0)^m$ . Quelques développements peuvent être faits pour traiter analytiquement quelques cas de défauts ou d'asphéricité et éviter l'utilisation des éléments finis. Wang [WB 03] détaille grâce à la théorie de propagation des rupture la probabilité de rupture SiC associée à une rupture du PyC, par des formules analytiques, qui est intégrée à TIMCOAT. Martin [Mar 02] traité analytiquement l'ovalisation des particules, et propose la simulation des couches cassées en mettant à zéro la rigidité de la couche concernée, ceci pouvant être fait en cours de calcul.

Miller [MW 94] traite des particules asphériques grâce à une simulation par EF. Pour retomber sur le calcul statistique que l'on peut avoir avec les formules analytiques, une particule

avec défaut sera représentée par une particule sphérique dont la contrainte (ou la containte à rupture) sera corrigée pour retomber sur la même probabilité. on veut trouver une transformation ou un  $\sigma'_0$  tel que

$$\int \left( \frac{\sigma_a}{\sigma_0} \right)^m dV = \int \left( \frac{\sigma_s}{\sigma'_0} \right)^m dV, \quad (4.2.11)$$

où  $\sigma_a$  est le champ de contrainte de la particule asphérique,  $\sigma_s$  la contrainte dans la particule sphérique et  $\sigma_0$  une caractéristique matériau.

en supposant que l'on peut écrire :

$$\int (\sigma_a)^m dV = \left( \int (\sigma_{an})^m dV \right) (f(t))^m \text{ et } \int (\sigma_s)^m dV = \left( \int (\sigma_{sn})^m dV \right) (g(t))^m,$$

où l'indice  $n$  signifie *nominal* (par exemple en fin d'irradiation).

la rupture dans une particule asphérique pourra être déterminée en comparant la contrainte à rupture à une contrainte équivalente  $\sigma$  écrite de la forme :

$$\sigma = \sigma_{ns} f(t) E(\varepsilon) \quad (4.2.12)$$

ou  $t$  est l'écart à la particule moyenne  $\varepsilon$  est l'asphéricité  $f$  et  $E$  sont des fonctions déterminées par des calculs EF. Finalement, une méthode de Monte Carlo directe est appliqué en tirant aléatoirement  $t$ ,  $\varepsilon$  et la contrainte à rupture.

Le code PARFUME [MPM 02] peut utiliser les EF pour simuler des particules non sphériques. Dans les cas usuels, la statistique est réalisée par Monte-Carlo direct à partir d'un modèle analytique simplifié prenant en compte la distribution des paramètres géométriques des particules. Il prend en compte de façon différentielle n'importe quel historique de température et fluence [MPVM 03].

Comme la modélisation des particules par EF est très longue et les probabilités recherchées très faibles, les résultats des simulations par EF seront toujours rattachées à des formules analytiques, sur lesquelles un calcul statistique sera réalisé.

### 4.3 Analyse de sensibilité du code ATLAS

Cette étude de sensibilité a fait l'objet d'une note technique [PC 05], nous en reprenons le contenu. Cette étude a pour objectif de mettre, tout d'abord, en évidence les sources d'incertitude en entrée et de décrire le comportement général du code soumis à cette incertitude, et aussi de classer les différentes sources d'incertitudes en deux catégories : paramètres influents ou non influents.

### 4.3.1 Incertitude en entrée

Les sources d'incertitude peuvent être classées en trois grandes catégories : les paramètres de fabrication des particules, les paramètres d'irradiation et les lois de comportement des matériaux.

Nous disposons d'assez peu de données concernant les variations des paramètres. En règle générale, elles sont suffisantes pour déterminer un minimum et un maximum. Ceci permet juste de déterminer une loi uniforme. Des données plus riches nous permettraient de considérer dans certains cas des lois normales. En l'absence d'information nous nous aiderons d'avis d'experts pour faire des hypothèses raisonnables.

#### 4.3.1.1 Les paramètres de fabrication

Les variations des caractéristiques des particules sont de nature statistique et dépendent du dispositif expérimental de fabrication, elles concernent la température de fabrication, les paramètres géométriques ainsi que la porosité ; ces variations sont assez bien connues de par les spécifications de fabrication ainsi que les contrôles qualité effectués tout au long du processus de fabrication (cf. tableau 4.3.1).

Variable	Description	Unité
TFAB	Température de fabrication	K
KERNRAD	Rayon initial du noyau	m
BUFFTHIC	Épaisseur initiale du buffer	m
BUFFPORI	Porosité initiale du buffer	/
BUFFOPR	Part initiale des porosités ouvertes du buffer	/
IPYCTHIC	Épaisseur initiale de IPyC	m
IPYCPORI	Porosité initiale du IPyC	/
SICTHIC	Épaisseur initiale du SiC	m
OPYCTHIC	Épaisseur initiale du OPyC	m
OPYCPORI	Porosité initiale du OPyC	/

TAB. 4.3.1 – Paramètres de fabrication des particules modélisant l'incertitude sur les paramètres de fabrication des particules

Les lois modélisant les paramètres de fabrication ont été déterminées à partir des documents [Hun 04], [Phe 05] et avis d'experts (cf. tableau 4.3.2).

#### 4.3.1.2 Les paramètres d'irradiation

Les particules sont réparties de façon aléatoire dans le coeur du réacteur ; les conditions extérieures auxquelles elles sont soumises sont variables et dépendent de leur position et de

Facteur	Lois normales tronquées					Facteur	Lois uniformes	
	min	m	max	$\sigma$	(cov)		min	max
KERNRAD ( $\mu\text{m}$ )	230	250	270	12,5	(5%)	TFAB (K)	1300	1600
BUFFTHIC ( $\mu\text{m}$ )	75	95	115	9,2	(10%)	BUFFPORI (%)	50	60
IPYCTHIC ( $\mu\text{m}$ )	30	40	50	2	(5%)	BUFFOPR (%)	80	100
SICTHIC ( $\mu\text{m}$ )	30	37	44	1,85	(5%)	IPYCPORI (%)	11	18
OPYCTHIC ( $\mu\text{m}$ )	30	40	50	2	(5%)	OPYCPORI (%)	11	18

TAB. 4.3.2 – *Lois de probabilité des paramètres de fabrication*

la conception du réacteur ; les distributions de ces quantités peuvent être déterminées par des calculs neutroniques et thermo-hydrauliques (cf. tableau 4.3.3) ; Un point doit être précisé concernant les paramètres réacteur. Le rapport appelé le facteur d'accélération (ACC) est caractéristique d'une position dans le réacteur et d'un type de réacteur, il est égal au rapport entre le flux neutronique nominal et la vitesse (ou probabilité) de fission nominale (PFN).

Variable	Description	Unité
PRESS	Pression (constante) appliquée à la couche externe de la particule	Pa
ACC	Facteur d'accélération	$\text{n.m}^{-2}$
PFN	Vitesse de fission nominale	$\text{s}^{-1}$
TIRR	Température d'irradiation	K

TAB. 4.3.3 – *Paramètres d'irradiation modélisant l'incertitude sur les conditions d'irradiation*

La détermination des lois de variations des paramètres d'irradiation dépend du type de réacteur ou de l'expérience choisie pour faire l'étude, nous présenterons cela au paragraphe 4.3.3.

#### 4.3.1.3 Lois de comportement

Les propriétés physiques et les lois de comportement des matériaux sont connues avec plus ou moins de précision ; l'incertitude sur ces lois est souvent difficile à approcher en pratique et repose essentiellement sur des avis d'experts. Nous avons choisi les lois BNFL pour l'étude (cf. tableau 4.3.4). Les couches IPyC et OPyC sont considérées comme un seul matériau, ce qui explique que ces deux couches partagent le même jeu de facteur.

L'incertitude sur les lois décrivant les propriétés des matériaux est difficile à déterminer. Certaines lois sont jugées plus fiables que d'autres. Nous choisissons de fixer la même incertitude sur toute les lois pour pouvoir les comparer entre elles. L'incertitude est de 10%, elle est introduite de la manière suivante : les lois de comportement sont multipliées par un facteur

Variable	Description	Unité
KERNNU	Coefficient de Poisson du noyau	/
KERNYOUN	Module d'Young du noyau	Pa
KERNDENS	Densification du noyau	/
KERNGONG	Gonflement gazeux du noyau	/
KERNGONS	Gonflement solide du noyau	/
KERNGARR	Taux de relachement des gaz de fission	/
KERNOPF	Quantité d'oxygène relachée pour un combustible	/
KERNLAMB	Conductivité thermique du noyau	W.m <sup>-1</sup> .K <sup>-1</sup>
KERNALPH	Dilatation thermique du noyau	K <sup>-1</sup>
BUFFNU	Coefficient de Poisson du buffer	/
BUFFYOUN	Module d'Young du buffer	Pa
BUFFKFLU	Coefficient de fluage du buffer	Pa <sup>-1</sup> .n <sup>-1</sup> .m <sup>2</sup>
BUFFDENS	Densification du buffer	/
BUFFLAMB	Conductivité thermique du buffer	W.m <sup>-1</sup> .K <sup>-1</sup>
BUFFALPH	Dilatation thermique du buffer	K <sup>-1</sup>
PYCNU	Coefficient de Poisson du PyC	/
PYCYOUN	Module d'Young du PyC	Pa
PYCKFLU	Coefficient de fluage du PyC	Pa <sup>-1</sup> .n <sup>-1</sup> .m <sup>2</sup>
PYCDENR	Densification radiale du PyC	/
PYCDENT	Densification orthoradiale du PyC	/
PYCLAMB	Conductivité thermique du PyC	W.m <sup>-1</sup> .K <sup>-1</sup>
PYCALPH	Dilatation thermique du PyC	K <sup>-1</sup>
SICNU	Coefficient de Poisson du SiC	/
SICYOUN	Module d'Young du SiC	Pa
SICKFLU	Coefficient de fluage du SiC	Pa <sup>-1</sup> .n <sup>-1</sup> .m <sup>2</sup>
SICDENS	Densification du SiC	/
SICLAMB	Conductivité thermique du SiC	W.m <sup>-1</sup> .K <sup>-1</sup>
SICALPH	Dilatation thermique du SiC	K <sup>-1</sup>

TAB. 4.3.4 – *Lois de comportement modélisant l'incertitude sur les lois de comportement*

qui varie selon une loi uniforme sur  $[0, 95; 1, 05]$ .

#### 4.3.2 Caractéristiques des réacteurs et expériences étudiés

L'analyse de sensibilité est réalisée sur deux configurations réacteur :

1. HFR-EU2 : le réacteur HFR est un réacteur expérimental à haut-flux situé à Petten aux Pays-Bas, EU2 est le nom d'une expérience prévue mais qui a été à finalement abandonnée ;
2. GT-MHR : c'est un projet de réacteur de production initié par les États-Unis (General Atomic) et la Russie ; D'autres partenaires les ont rejoint, dont AREVA.

La figure 4.2.3 représente l'historique de la température et de la fluence, la vitesse de fission est considérée proportionnelle à la fluence rapide et la pression extérieure est supposée constante au cours du temps.

Le tableau 4.3.5 présente les conditions d'irradiation de ces deux réacteurs [FLM<sup>+</sup> 05, Lan 02, LSSQ 02, CLRH 01, Fab 01]. Les distributions des paramètres d'irradiation sont détaillées dans le tableau 4.3.6.

Paramètres d'irradiation	HFR-EU2	GT-MHR
Pression caloporteur (bar)	4.5	70
Température min (K)	1323	775
Température max (K)	1423	1175
Taux de combustion max (%FIMA)	10	13,7
Flux rapide max. ( $n.m^{-2}$ )	4,510 <sup>25</sup>	3,510 <sup>25</sup>
Durée d'irradiation (JEPP)	350	490
JEPP : jour équivalent pleine puissance		
Flux neutronique nominal ( $n.m^{-2}.s^{-1}$ )	1,4810 <sup>18</sup>	8,2810 <sup>17</sup>
= $\frac{\text{flux rapide max.}}{\text{durée d'irradiation (s)}}$		
Proba de fission nominale ( $s^{-1}$ )	3,3010 <sup>-9</sup>	3,2310 <sup>-9</sup>
= $\frac{\text{taux de combustion max.}}{\text{durée d'irradiation (s)}}$		

TAB. 4.3.5 – Conditions d'irradiation des réacteurs HFR et GTM-HR

Facteur	HFR-EU2			GT-MHR		
	valeur nominale	Lois uniformes		valeur nominale	Lois uniformes	
		min	max		min	max
PRESS (bar)	5	1	5	70	60	70
ACC ( $n.m^{-2}$ )	4,4810 <sup>26</sup>	4,0510 <sup>26</sup>	4,9510 <sup>26</sup>	2,5610 <sup>26</sup>	2,3010 <sup>26</sup>	2,8110 <sup>26</sup>
PFN ( $s^{-1}$ )	3,3010 <sup>-9</sup>	2,2010 <sup>-9</sup>	6,6110 <sup>-9</sup>	3,2310 <sup>-9</sup>	2,1510 <sup>-9</sup>	6,4710 <sup>-9</sup>
TIRR (K)	1373	1323	1423	975	775	1175

TAB. 4.3.6 – Lois de probabilité des lois de comportement, la valeur nominale n'est pas systématiquement le milieu de l'intervalle. La durée d'irradiation et le taux de combustion maximum est de 350 JEPP (jour équivalent pleine puissance) et 10% FIMA pour HFR-EU2 et de 490 JEPP et 13.5% FIMA pour GTM-HR

### 4.3.3 Résultats

Nous présentons ici les résultats obtenus sur les deux configurations réacteurs. Les méthodes d'analyse de sensibilité sont décrites en annexe A.

La rupture mécanique d'une particule peut être provoquée par la rupture d'une des couches denses externes (IPyC, SiC, OPyC). De plus, nous avons vu au paragraphe 4.1.2.3 que la rupture des couche OPyC et IPyC favorise considérablement la rupture de la couche de SiC ayant pour conséquence un relâchement très important des produits de fission hors de la particule. Les réponses du code ATLAS que nous allons donc étudier sont les contraintes orthoradiales maximales dans ces couches IPyC, OPyC et SiC. Ces maxima se situent en début d'irradiation pour les couches IPyC et OPyC et en fin d'irradiation pour la couche SiC (voir figure 4.1.5).

#### 4.3.3.1 HFR-EU2

Une première analyse d'incertitude permet de mettre en évidence la grande variabilité de la contrainte orthoradiale de la couche SiC, et la faible variabilité des contraintes orthoradiales des autres couches.

La figure 4.3.1 présente les statistiques élémentaires sur les lois des trois réponses du Code, les résultats ont été obtenus en ne faisant varier que les paramètres de fabrication et les paramètres d'irradiation. Nous constatons que le coefficient de variation est très faible pour les couches PyC (moins de 2 %), et plus important pour la couche SiC (26%). L'incertitude sur la contrainte se situe donc au niveau de la couche SiC. En faisant varier les paramètres d'irradiation, nous obtenons des résultats similaires.

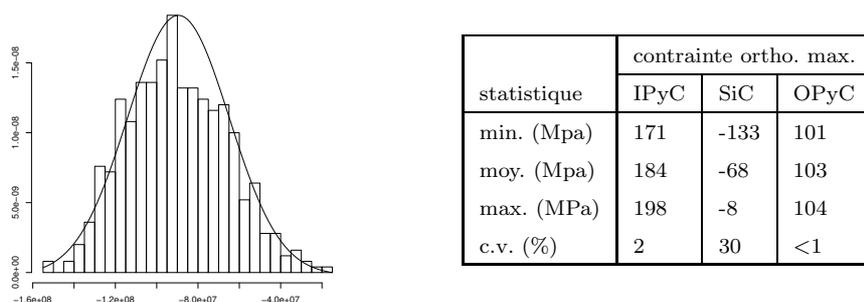


FIG. 4.3.1 – HFR-EU2 - Analyse d'incertitude liée aux paramètres de fabrication et d'irradiation. L'histogramme représente la distribution de la seconde réponse (contrainte dans le SiC); nous avons superposé la densité de la loi normale estimée. La tableau regroupe les statistiques élémentaires pour les trois réponses.

Nous venons de voir que seule la contrainte dans la couche SiC est soumise à une incertitude. Nous allons donc tenter d'expliquer cette incertitude en faisant une analyse de sensibilité. L'objectif est de déterminer les paramètres qui sont responsables de cette incertitude. Pour

cela, nous utilisons un modèle linéaire, la régression linéaire du modèle est de très bonne qualité (que se soit en faisant varier les paramètres de fabrication seuls ou avec les paramètres d'irradiation). Les résultats de l'analyse de sensibilité sont présentés dans le tableau 4.3.7.

Remarquons que nous effectuons une analyse de sensibilité globale, nous nous intéressons à la variabilité de la sortie du modèle dans son domaine de variation, c'est-à-dire que nous étudions comment la variabilité des paramètres entrées se répercute sur celle de la sortie, en déterminant quelle part de variance de la sortie est due à tels paramètres entrées ou tel ensemble de paramètres d'entrées.

Facteur	PEAR	SRC
TFAB	-0.13	0.0
KERNRAD	0.51	0.26
BUFFTHIC	-0.52	0.24
IPYCTHIC	-0.1	0.03
SICTHIC	0.24	0.03
OPYCTHIC	-0.23	0.06
BUFFPORI	-0.26	0.08
BUFFOPR	-0.14	0.02
IPYCPORI	0.05	0.0
OPYCPORI	-0.02	0.0
PRESS	-0.02	0.0
ACC	-0.38	0.11
PFN	0.07	0.0
TIRR	0.33	0.13

TAB. 4.3.7 – Réacteur HFR-EU2 : Analyse de sensibilité liée aux paramètres de fabrication et aux paramètres d'irradiation. Les indices SRC quantifient en % l'influence d'un paramètre et le signe des indices PEAR permet de connaître le sens de l'influence.

Seulement quatre paramètres sont responsables de plus de 70% de la variation de la contrainte : le rayon du noyau, l'épaisseur du buffer, la température d'irradiation et le facteur d'accélération.

Concernant les paramètres de fabrication,

- la contrainte augmente quand le rayon du noyau augmente ; en effet, un plus gros noyau représente plus de combustible et donc plus de gaz de fission relâché lors de l'irradiation ;
- la contrainte diminue quand l'épaisseur du buffer augmente : en effet un plus gros buffer représente plus d'espace disponible pour les gaz de fission relâchés ;
- la contrainte diminue quand la porosité du buffer augmente : la raison est la même que la précédente.

Les trois facteurs responsables des variations de la contrainte dans le SiC sont donc liés au relâchement des gaz. Nous pouvons donc penser que c'est la variabilité de ce phénomène qui explique quasiment à lui seul les importantes variations de la contrainte dans la couche SiC.

L'influence des conditions d'irradiation est importante (25%) et passe par la température d'irradiation et le facteur d'accélération. Ce résultat est à nuancer car les données sur ces paramètres sont légères. Nous pouvons seulement dire que ces deux paramètres sont potentiellement influents. Une meilleure connaissance des intervalles de variation de ces paramètres permettrait d'affiner ce résultat.

#### 4.3.3.2 GT-MHR

Comme pour HFR-EU2 une première analyse d'incertitude permet de mettre en évidence la grande variabilité de la contrainte orthoradiale de la couche SiC, et la faible variabilité des contraintes orthoradiales des autres couches (cf. figure 4.3.2).

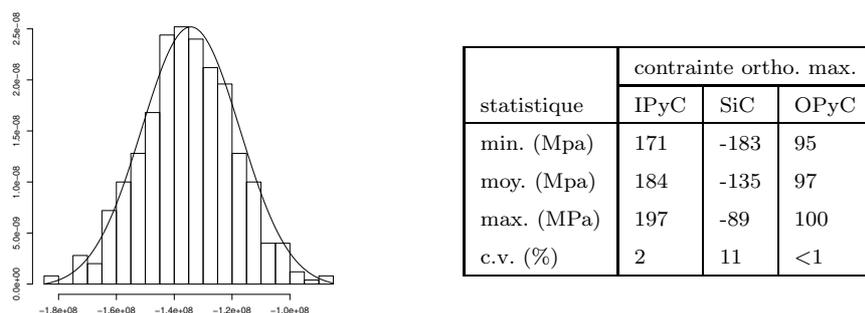


FIG. 4.3.2 – GT-MHR - Analyse d'incertitude liée aux paramètres de fabrication et d'irradiation. L'histogramme représente la distribution de la seconde réponse (contrainte dans le SiC); nous avons superposé la densité de la loi normale estimée. La tableau regroupe les statistiques élémentaires pour les trois réponses.

Les résultats de l'analyse de sensibilité sont présentés dans le tableau 4.3.8. Seulement quatre paramètres sont responsables de plus de 65% de la variation de la contrainte : la température d'irradiation, l'épaisseur du buffer, l'épaisseur du SiC et le rayon du noyau. Par rapport à l'étude précédente sur HFR-EU2 :

- le facteur d'accélération n'est plus influent ;
- l'épaisseur du SiC devient influente; la différence s'explique par les conditions d'irradiation. À température élevée, la variabilité sur l'épaisseur du buffer est plus influente que celle du SiC sur la valeur finale de la contrainte. À température plus basse, la variabilité

sur l'épaisseur du SiC est plus influente que celle du buffer, mais plus le SiC est épais et moins sa contrainte est négative.

Facteur	PEAR	SRC
TFAB	0.02	0.0
KERNRAD	0.37	0.12
BUFFTHIC	-0.39	0.17
IPYCTHIC	-0.29	0.07
SICTHIC	0.43	0.14
OPYCTHIC	-0.3	0.06
BUFFPORI	-0.21	0.04
BUFFOPR	-0.02	0.01
IPYCPORI	0.04	0.0
OPYCPORI	0.1	0.0
PRESS	-0.11	0.01
ACC	-0.09	0.01
PFN	0.17	0.02
TIRR	0.56	0.29

TAB. 4.3.8 – Réacteur GT-MHR : Analyse de sensibilité liée aux paramètres de fabrication et aux paramètres d'irradiation. Les indices SRC quantifient l'influence en % d'un paramètre et le signe des indices PEAR permet de connaître le sens de l'influence

**Remarque :** Pour étudier l'influence de l'ensemble des paramètres d'entrées, ceux de fabrication, irradiation et lois de comportement (que nous n'avons pas pris en compte jusqu'à présent), nous avons choisi d'utiliser une méthode de screening (la méthode de MORRIS présentée en annexe A) qui analyse qualitativement l'importance des paramètres d'entrées sur la variabilité de la réponse du modèle étudiée. Cette méthode est plus adaptée pour déterminer si potentiellement une loi de comportement peut influencer sur la réponse car l'incertitude des lois est arbitraire par rapport aux incertitudes sur les autres paramètres, cette influence sera alors effective si l'incertitude sur la loi est réellement importante. Les résultats de cette étude sont présentés sur la figure 4.3.3. Nous constatons que :

- la contrainte orthoradiale dans la couche SiC est sensible majoritairement aux paramètres géométriques (épaisseur des couches) ainsi qu'à la température d'irradiation ; la contrainte dans le SiC est aussi sensible aux variations de l'épaisseur des couches IPyC, SiC et OPyC. Ces variations sont faibles ce qui explique que nous n'ayons pas constaté dans l'analyse de sensibilité précédente la forte influence de ces paramètres (sauf l'épaisseur du SiC) ;
- en ce qui concerne les lois de comportement, seuls le coefficient de fluage du PyC et la

densification orthoradiale du PyC sont des paramètres auxquels la contrainte orthoradiale du SiC est sensible.

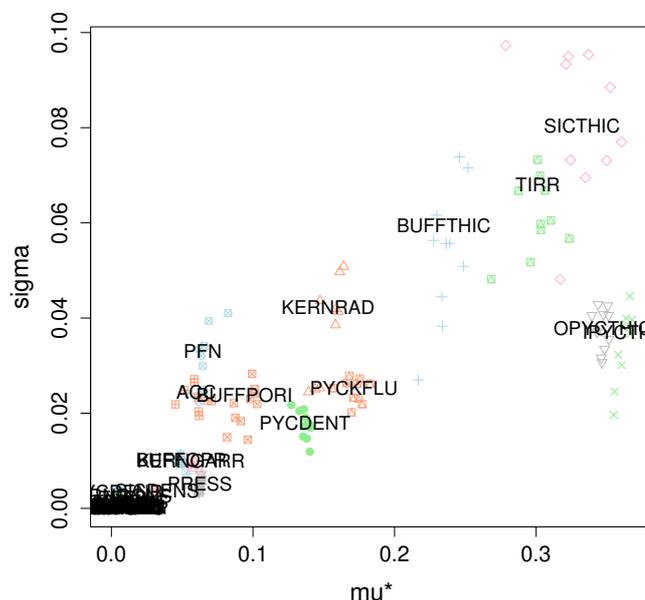


FIG. 4.3.3 – Réacteur GT-MHR : Analyse de sensibilité des paramètres de fabrication, d'irradiation et des lois de comportement. Les estimations des mesures d'influence de MORRIS sont représentées par un graphe  $(\mu^*, \sigma)$ . L'axe  $\mu^*$  représente la sensibilité. L'axe  $\sigma$  représente les interactions ou les effets non linéaires.

## 4.4 Fiabilité du combustible

Nous avons utilisés les méthodes du tirage d'importance pour estimer la probabilité de rupture du combustible à particules sur l'expérience HFR-EU1 et des réacteurs de type GT-MHR et PBMR.

### 4.4.1 HFR-EU1

Le réacteur HFR est un réacteur expérimental à haut-flux, EU1 est le nom de l'expérience. Le taux de combustion des particules est de l'ordre de 20% FIMA, et la fluence rapide de l'ordre de  $6.10^{25}$  neutrons.m<sup>-2</sup>.s<sup>-1</sup>.

Les couches sont fabriquées à haute température et nous les supposons sans contrainte et au contact avec le noyau, lors du retour à température ambiante, un jeu entre le noyau et les couches s'est créé ainsi que les contraintes dans les couches. Par simplicité, nous démarrons la

simulation à la fin de la fabrication. A partir du moment où les particules sont fabriquées pour être dans le réacteur, nous définissons les cinq phases de variation des conditions que subissent les particules :

1. la température décroît de  $1500K$  jusqu'à la température ambiante ( $300K$ ) en 10000 secondes ;
2. la température croît jusqu'à sa valeur nominale ( $1223K$  ou  $1373 K$ ) en 10000 secondes avec un flux neutronique et une vitesse de fission croissants ;
3. les conditions de température ne changent pas pendant les 600 jours d'irradiation ; pendant cette période, la particule atteint linéairement son taux de combustion et sa fluence finale ;
4. la température décroît ensuite jusqu'à  $300K$  sans flux neutronique ;
5. essai de chauffage de  $1500K$  à  $1900K$  (cette phase ne sera pas prise en compte dans l'évaluation de la probabilité de rupture).

La figure 4.4.1 illustre les résultats obtenus avec ATLAS. Nous avons utilisé les lois BNFL pour les paramètres de fabrication.

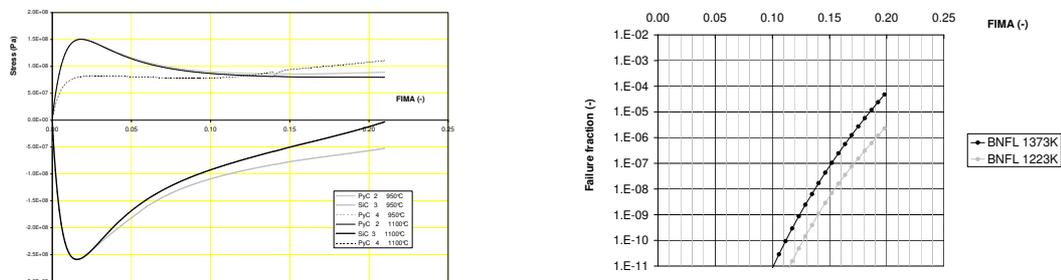


FIG. 4.4.1 – *HFR-EU1 - Résultats. A gauche : Évolution des contraintes orthoradiale dans les couches IPyC, SiC et OPyC en fonction du taux de combustion. A droite : Évolution de la probabilité de rupture en fonction du taux de combustion.*

Nous utilisons une méthode de tirage d'importance avec des densités d'importance prédéfinies par jugement d'expert (cf. tableau 4.4.1).

Pour une température d'irradiation de  $1373K$ , la probabilité de rupture au dernier pas de temps est  $2.2510^{-5}$  et sa variance :  $2.110^{-11}$ .

Pour une température d'irradiation de  $1223K$ , la probabilité de rupture au dernier pas de temps est  $4.210^{-6}$  et sa variance :  $7.810^{-13}$ .

	Diamètre Noyau	Épaisseur Buffer	Épaisseur IPyC	Épaisseur SiC	Épaisseur OPyC
pdf initiale	$\mathcal{N}(502.2,10.8)$	$\mathcal{N}(94.9,14.3)$	$\mathcal{N}(41,3.4)$	$\mathcal{N}(35.4,1.9)$	$\mathcal{N}(40,3.8)$
pdf d'importance	$\mathcal{N}(502.2,10.8)$	$\mathcal{N}(70,28.6)$	$\mathcal{N}(41,6.8)$	$\mathcal{N}(35.4,3.8)$	$\mathcal{N}(40,7.6)$

TAB. 4.4.1 – Réacteur HFR-EU1 : densités initiales et d'importance de l'épaisseur des couches de la particule (jugement d'expert).

A partir de 2000 simulations du code ATLAS, nous établissons un polynôme de degré 3 prédisant la contrainte orthoradiale de la couche SiC en fin d'irradiation. Sur ce polynôme, nous pouvons vérifier la méthode du tirage d'importance à plusieurs étapes (cf. chapitre 2). Les densités initiales, intermédiaires et d'importance sont illustrées sur la figure 4.4.2.

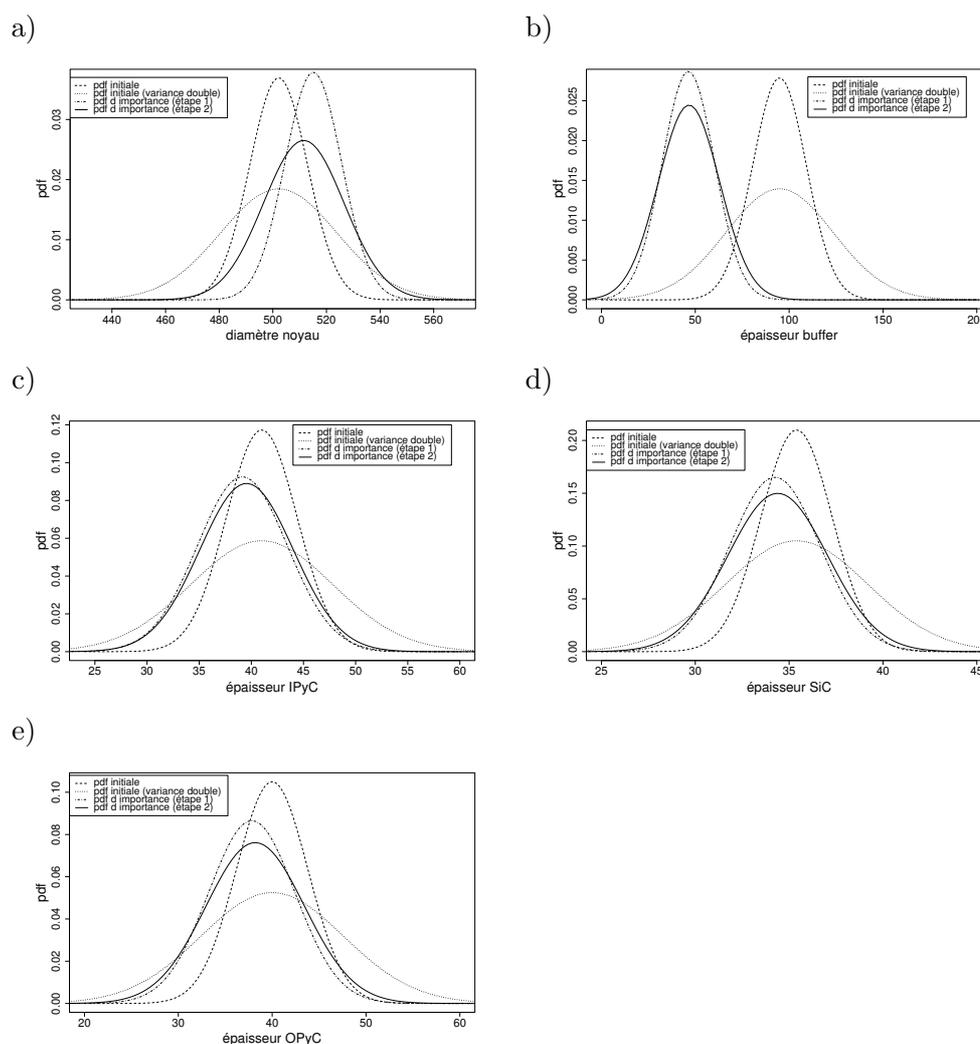


FIG. 4.4.2 – HFR-EU1 (polynôme) Densités initiale et densités d'importance de l'épaisseur des couches de la particule : a) noyau, b) buffer, c) IPyC, d) SiC, e) OPyC

La figure 4.4.3 représente les résultats obtenus par une méthode de tirage d'importance à 3 étapes (cf § 2.3.1.1). Le tableau 4.4.2 récapitule quelques résultats obtenus en faisant varier le nombre de simulation total des trois étapes (une augmentation du nombre d'étapes n'améliore pas les résultats pour ce polynôme).

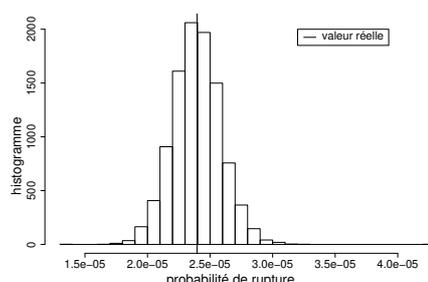


FIG. 4.4.3 – Histogramme des probabilités de rupture (polynôme) obtenues par le tirage d'importance à 3 étapes (200,200,200). Les histogrammes sont tracés à partir de 10000 expériences.

Nombre de simulations	100 100	100 100	200 200	300 300	300 300	400 400	300 300
	100	200	200	200	400	400	600
Moyenne	$2.39 \cdot 10^{-5}$						
Coeff. var.	21.6%	15.5%	8.1%	6.7%	5.7%	5.5%	4.5%

TAB. 4.4.2 – Comparaison du coefficient de variation des probabilités de rupture obtenues par le tirage d'importance à 3 étapes en fonction du nombre de simulations sur le polynôme. Les résultats sont obtenus à partir de 10000 expériences.

## 4.4.2 PBMR

Le PBMR (Pebble Bed Modular Reactor) est un projet de réacteur HTR, en Afrique, de puissance thermique 400MW [Rei 04]. Nous nous intéressons à son design.

### 4.4.2.1 Gestion du combustible

Dans ce type de réacteur, l'élément combustible est un boulet contenant 9g d'uranium. Le coeur est un cylindre de rayon interne 1m, de rayon externe 1.85m et de hauteur 11m. Il contient 450000 boulets, ce qui donne un total d'environ 3.5 milliards de particules combustible. Les boulets sont poussés par le haut du coeur et les boulets sortant (par le bas) sont alors replacés en haut du coeur. La vitesse de passage dépend de la position du boulet dans le coeur. Le temps

de chute est en moyenne égale à 154.7 jours, ce qui donne un taux nominal de combustion de 9.7% FIMA après 6 passages.

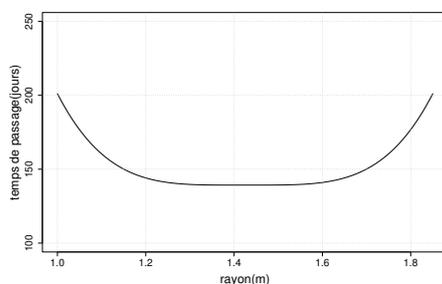


FIG. 4.4.4 – Temps de passage des boulets en fonction de leur place dans le coeur

La position radiale de l'élément combustible est prise aléatoirement dans le réacteur avec un rayon au carré de densité uniforme  $\mathcal{U}(1, 3.4225)$ . La position de la particule dans l'élément combustible est prise aléatoirement avec un rayon au cube de loi uniforme  $\mathcal{U}(0, 1.562410^{-5})$ .

#### 4.4.2.2 Conditions en réacteur

Compte tenu de la puissance du réacteur et de la quantité d'uranium dans le coeur, la moyenne de vitesse de fission doit être de  $1.21 \cdot 10^{-9} \text{s}^{-1}$  et nous prenons une fluence rapide moyenne égale à  $6 \cdot 10^{17} \text{neutrons.m}^{-2}.\text{s}^{-1}$ . Ces paramètres d'irradiation dépendent eux aussi de la position dans le cylindre, nous leur affectons un facteur correctif en fonction de leur place dans le cylindre (cf. figure 4.4.5).

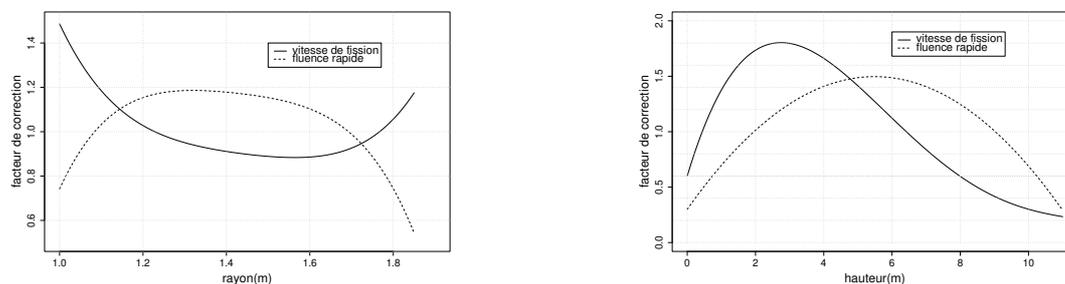


FIG. 4.4.5 – Facteur correctif de la vitesse de fission et de la fluence rapide en fonction de la place du boulet dans le cylindre, à gauche le rayon, à droite la hauteur

Nous affectons aussi un facteur correctif à la vitesse de fission qui dépend du nombre de passages des boulets :  $1.5 - 0.6 \frac{i}{5}$  où  $i$  est le nombre de passages. Après 6 passages, le boulet est enlevé du coeur.

La température moyenne du caloporteur en entrée (haut du cylindre) est de  $500^{\circ}$  C et la température moyenne du caloporteur en sortie (bas du cylindre) est de  $950^{\circ}$  C. La température d'irradiation dépend aussi de la position de la particule et l'expression de la température d'irradiation d'une particule est exprimée en fonction de sa position, du nombre de passages du boulet, et de la température du caloporteur :

$$T_{part} = T_{He} + \frac{Q}{4\lambda_p\pi} \left[ \frac{1}{r_{int}} - \frac{1}{r_{ext}} + \frac{1}{2r_{int}^3} (r_{int}^2 - r^2) \right] + \frac{2Qr_{ext}}{A\nu\lambda_{He}}, \quad (4.4.1)$$

où :

- $T_{He}$  la température du caloporteur :

$$T_{He} = \left( 38.65 \int \text{CorV}_1 dz - 500 \right) (\text{CorV}_2(1 - z/13.75) + z/13.75) + 500,$$

où  $\text{CorV}_1$  et  $\text{CorV}_2$  sont les facteurs correctifs de la vitesse de fission respectivement en fonction du rayon et de la hauteur ;

- $r$  la position radiale de la particule dans le boulet ;
- $r_{int} = 2.5\text{cm}$  rayon de l'intérieur du boulet, rempli avec les particules ;
- $r_{int} = 3\text{cm}$  rayon total du boulet ;
- $A$  la surface d'un boulet ;
- $\lambda_p$  la conductivité du boulet ;
- $\lambda_{He} = 0.3$  la conductivité du caloporteur hélium ;
- $\nu = 353$  ;
- $Q$  la puissance totale du boulet.

Après le sixième passage, nous élevons la température à  $1873$  K pour représenter une situation accidentelle, c'est à ce moment là que nous évaluons la fiabilité. Le taux de combustion final n'est qu'en moyenne  $8.33\%$  FIMA.

#### 4.4.2.3 Résultats

Le tableau 4.4.3 récapitule les paramètres aléatoires en entrée du code ATLAS.

La figure 4.4.6 illustre l'évolution des contraintes orthoradiales dans les couches d'IPyC, de SiC et d'OPyc au cours du temps ; ces résultats ont été obtenu avec ATLAS.

Pour estimer la probabilité de rupture nous avons utilisé le tirage d'importance adaptatif présenté dans le chapitre 2. Nous avons choisi une stratégie à trois étapes :  $n_1 = 200$ ,  $n_2 = 200$  et  $n_3 = 400$ . Les résultats sont décrits dans le tableau 4.4.4 et illustrés sur la figure 4.4.7.

L'estimation de la probabilité de rupture n'est pas assez précise (coefficient de variation de

Paramètres aléatoires	Lois
Diamètre du noyau	$\mathcal{N}(251 \cdot 10^{-6}, 5.4 \cdot 10^{-6})$
Épaisseur du buffer	$\mathcal{N}(94.9 \cdot 10^{-6}, 14.3 \cdot 10^{-6})$ tronqué à $[5. \cdot 10^{-6}, 200. \cdot 10^{-6}]$
Épaisseur de IPyC	$\mathcal{N}(41.0 \cdot 10^{-6}, 3.4 \cdot 10^{-6})$
Épaisseur du SiC	$\mathcal{N}(35.4 \cdot 10^{-6}, 1.9 \cdot 10^{-6})$
Épaisseur de OPyC	$\mathcal{N}(40.0 \cdot 10^{-6}, 3.8 \cdot 10^{-6})$
Carré de la position radiale du boulet	$\mathcal{U}(1, 3.4225)$
Cube de la position radiale de la particule dans le boulet	$\mathcal{U}(0, 1.5625 \cdot 10^{-5})$

TAB. 4.4.3 – PBMR : Paramètres aléatoires (pour les gaussiennes, c'est l'écart-type qui est indiqué et non la variance)

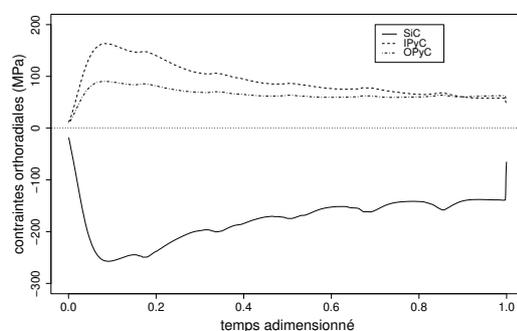


FIG. 4.4.6 – Évolution des contraintes orthoradiales moyennes dans les couches d'IPyC, SiC et OPyC en fonction du temps

	probabilité de rupture	écart-type estimé	écart-typ estimé par bootstrap
estimateur sans recyclage (2.3.1.1)	$1.0116 \cdot 10^{-7}$	$4.17756 \cdot 10^{-8}$	$4.1812 \cdot 10^{-8}$
estimateur avec recyclage (2.3.1.2)	$5.4925 \cdot 10^{-7}$	$3.49 \cdot 10^{-7}$	$3.492 \cdot 10^{-7}$
estimateur mixte (§ 2.4.3.4)	$4.063 \cdot 10^{-7}$		$1.0491 \cdot 10^{-7}$

TAB. 4.4.4 – PBMR - Probabilité de rupture , premier essai ( $n_1 = 200$ ,  $n_2 = 200$  et  $n_3 = 400$ )

l'ordre de 40% pour l'estimateur sans recyclage), il serait intéressant d'utiliser un nombre de simulations plus importants.

Nous optons encore pour une stratégie à trois étapes :  $n_1 = 200$ ,  $n_2 = 200$  et  $n_3 = 1000$ . Les résultats sont décrits dans le tableau 4.4.5 et illustrés sur la figure 4.4.8. L'estimation de la probabilité de rupture gagne un peu en précision (coefficient de variation de l'ordre de 30%).

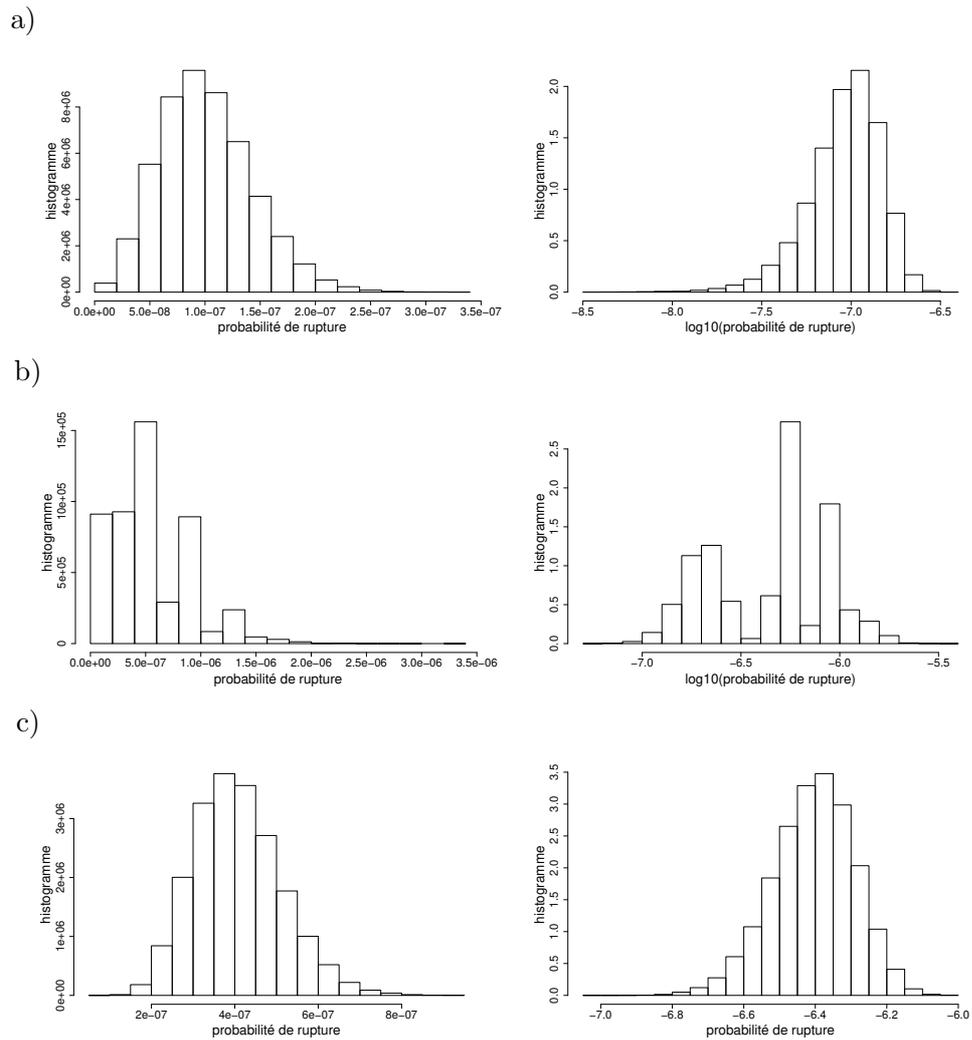


FIG. 4.4.7 – Histogramme de la probabilité de rupture estimée obtenu par bootstrap sur 50000 ré-échantillonnages - PBMR, premier essai ( $n_1 = 200$ ,  $n_2 = 200$  et  $n_3 = 400$ ). a) estimateur sans recyclage (2.3.1.1), b) estimateur avec recyclage (2.3.1.2), c) estimateur mixte (§ 2.4.3.4)

	probabilité de rupture	écart-type estimé	écart-typ estimé par bootstrap
estimateur sans recyclage (2.3.1.1)	$1.17485 \cdot 10^{-7}$	$3.6971 \cdot 10^{-8}$	$3.6922 \cdot 10^{-8}$

TAB. 4.4.5 – PBMR - Probabilité de rupture, deuxième essai ( $n_1 = 200$ ,  $n_2 = 200$  et  $n_3 = 1000$ )

### 4.4.3 GT-MHR

Les conditions d'irradiation d'un réacteur de type GT-MHR (ou ANTARES pour ce qui concerne le réacteur de AREVA-NP), n'étant pas publiées, nous allons beaucoup utiliser la description du réacteur PBMR vue au paragraphe précédent. Les performances du combustible

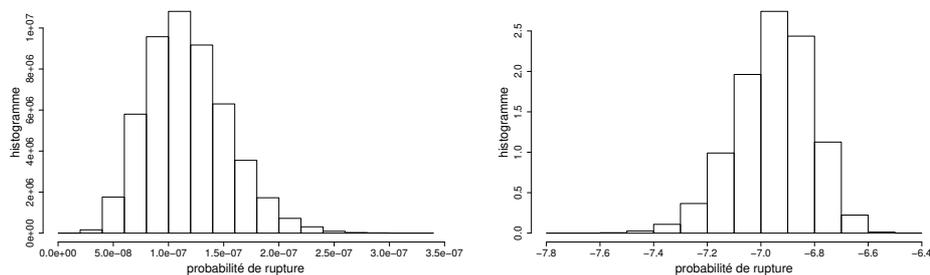


FIG. 4.4.8 – Histogramme de la probabilité de rupture estimée obtenu par bootstrap sur 50000 ré-échantillonnages - PBMR, deuxième essai ( $n_1 = 200$ ,  $n_2 = 200$  et  $n_3 = 1000$ ). Estimateur sans recyclage (2.3.1.1)

ne seront donc pas forcément représentatives.

#### 4.4.3.1 Gestion du combustible

Dans ce type de réacteur, l'élément combustible est un compact et a une place fixe durant un cycle d'irradiation. Parmi les gestions possibles un combustible, nous considérons que le combustible sera irradié en deux cycles avant d'être mis au rebut. Entre ses deux passages, le combustible verra sa position verticale modifiée et sa position radiale conservée.

Comme la vitesse de combustion varie négativement beaucoup en fonction de la température, nous utilisons la gestion suivante : le combustible est placé en partie basse chaude lorsqu'il est neuf, puis en position haute plus froide pour son deuxième cycle. Sur les schémas industriels, les combustibles vont plutôt, neuf, aux extrémités puis, usagés, au centre, mais ce que nous considérons convient mieux pour les conditions simples que nous prenons. La différence vient peut être de l'influence des barres de contrôle.

la première position verticale (distance au sommet) est aléatoire et suit une loi uniforme  $\mathcal{U}(5.5, 11)$ . Lors du second passage, les éléments sont rapprochés du haut d'une mi-hauteur.

Bien que le cœur soit formé d'assemblages hexagonaux, nous l'approchons par un cylindre troué. La position de l'élément combustible est prise aléatoirement dans le réacteur avec un rayon au carré de densité uniforme  $\mathcal{U}(1, 3.4225)$ .

#### 4.4.3.2 Conditions en réacteur

Nous supposons que le réacteur a une puissance totale, une vitesse de fission moyenne et une géométrie similaire à celle du réacteur PBMR. De par la géométrie et la disposition, très différentes, la densité en combustible est différente dans le coeur et donc tous ces paramètres ne

peuvent être conservés. Cependant, nous supposons que pour une puissance totale identique, c'est la taille du coeur qui sera différente, mais que les paramètres concernant les particules restent corrects moyennant un changement de variable sur les paramètres de position.

Dans un réacteur de type GT-MHR, il est possible de disposer le combustible avec un enrichissement variable en fonction de la position et la quantité de graphite dans le coeur est plus importante. Cela permet d'obtenir une vitesse de fission à peu près identique pour tout rayon. La position de certains objets dans le coeur (barres de contrôle par exemple), et le nombre réduit d'enrichissements différents utilisés font qu'il reste une variation, mais celle-ci est plus faible et supposée découplée des autres paramètres. Pour le facteur correctif de la vitesse de fission en fonction du rayon, nous prendrons alors une gaussienne  $\mathcal{N}(1, 0.1)$  tronquée à  $[0.5, 1.5]$ .

Nous négligeons l'effet de la gestion du combustible sur l'évolution verticale de la fluence rapide et de la vitesse de fission et nous conservons les mêmes profils que pour le réacteur PBMR (les mêmes facteurs correctifs). En absence de données sur ce sujet, nous gardons aussi le même profil radial de la fluence rapide. Nous modifions le facteur correctif de la vitesse de fission en fonction du nombre de passages pour ne prendre en compte que deux états :  $1.6 - 0.4i$  où  $i$  est le nombre de passages. Enfin, nous supposons que le caloporteur a une température identique au PBMR en entrée et en sortie, ceci n'étant pas fixé par le coeur. Comme nous avons le même profil vertical de densité de fission, nous conservons la même évolution verticale moyenne de la température du caloporteur que pour le réacteur PBMR. Comme les éléments sont souvent verticaux et traversant le réacteur, la fluctuation radiale de la densité de fission sera prise en compte pour la température du caloporteur, celui-ci ne subissant aucun « remélange ». La particule est positionnée aléatoirement dans le compact dont la température a une forme parabolique. Il y a ensuite des gradients entre le bord du compact et le bloc en graphite dû à un jeu gazeux et à un gradient dans le bloc graphite. Finalement, nous considérons la température d'une particule suivante :

$$T_{part} = T_{He} + Q \left( \frac{(r_{ext}^2 - r^2)}{4.V.\lambda_{compact}} + \frac{l_{gas}}{S.\lambda_{gas}} + \frac{l_{graphite}}{S.\lambda_{graphite}} \right), \quad (4.4.2)$$

où :

•  $T_{He}$  la température du caloporteur :

$$T_{He} = 34.3587 \left( \left( \left( \left( -\frac{0.000686}{5} Z + \frac{0.023688}{4} \right) Z - \frac{0.274258}{3} \right) Z + \frac{1.026548}{2} \right) Z + 0.600642 \right) Z \text{corF}_1 + 773,$$

où  $\text{CorV}_1$  facteur correctifs de la vitesse de fission en fonction du rayon ;

- $S = 1.92 \cdot 10^{-3} \text{ (m}^2\text{)}$  surface latérale du compact ;
- $V = 6.01 \cdot 10^{-6} \text{ (m}^3\text{)}$  le volume du compact ;
- $r_{ext} = 6.25 \cdot 10^{-3} \text{ (m)}$  le rayon du compact ;
- $r$  la position de la particule dans le compact ;
- $\lambda_{graphite} = \lambda_{compact} = 20 \text{ (W/m/K)}$  ;
- $l_{gas} = 25 \cdot 10^{-6} \text{ (m)}$  ;
- $l_{graphite} = 2.2 \cdot 10^{-3} \text{ (m)}$ , c'est un chemin équivalent ;
- $Q$  la puissance totale du compact.

Après les deux cycles, nous élevons la température à 1873 K pour représenter une situation accidentelle, c'est à ce moment là que nous évaluons la fiabilité. Le taux de combustion final est en moyenne 9.56% FIMA.

#### 4.4.3.3 Résultats

Le tableau 4.4.6 récapitule les paramètres aléatoires en entrée du code ATLAS.

Paramètres aléatoires	Lois
Diamètre du noyau	$\mathcal{N}(251 \cdot 10^{-6}, 5.4 \cdot 10^{-6})$
Épaisseur du buffer	$\mathcal{N}(94.9 \cdot 10^{-6}, 14.3 \cdot 10^{-6})$ tronqué à $[5 \cdot 10^{-6}, 200 \cdot 10^{-6}]$
Épaisseur de IPyC	$\mathcal{N}(41.0 \cdot 10^{-6}, 3.4 \cdot 10^{-6})$
Épaisseur du SiC	$\mathcal{N}(35.4 \cdot 10^{-6}, 1.9 \cdot 10^{-6})$
Épaisseur de OPyC	$\mathcal{N}(40.0 \cdot 10^{-6}, 3.8 \cdot 10^{-6})$
Position radiale du compact	$\mathcal{U}(1, 3.4225)$
Position verticale du compact en début de vie	$\mathcal{U}(5.5, 11)$
Carré de la position radiale de la particule dans le compact	$\mathcal{U}(1, 3.4225)$
facteur correctif vitesse de fission	$\mathcal{N}(1, 0.1)$ tronquée $[0.5, 1.5]$

TAB. 4.4.6 – *GT-MHR : Paramètres aléatoires (pour les gaussiennes, c'est l'écart-type qui est indiqué et non la variance)*

La figure 4.4.9 illustre l'évolution des contraintes orthoradiales dans les couches d'IPyC, de SiC et d'OPyC au cours du temps ; ces résultats ont été obtenu avec ATLAS.

Pour estimer la probabilité de rupture nous avons utilisé le tirage d'importance adaptatif présenté dans le chapitre 2. Nous avons choisi une stratégie à trois étapes :  $n_1 = 200$ ,  $n_2 = 200$  et  $n_3 = 200$ . Les résultats sont décrits dans le tableau 4.4.7)

L'estimation de la probabilité de rupture est précise (coefficient de variation de l'ordre de 7% pour l'estimateur sans recyclage).

**Remarque :** Bien que la probabilité de rupture soit plus élevée ici que sur la concept précédent,

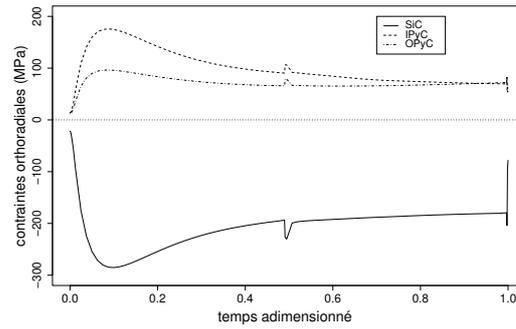


FIG. 4.4.9 – Évolution des contraintes orthoradiales dans les couches d'IPyC, SiC et OPyC en fonction du temps. Les pics sont obtenus lors de l'inter-cycle.

	probabilité de rupture	écart-type estimé	écart-typ estimé par bootstrap
estimateur sans recyclage (2.3.1.1)	$1.11558 \cdot 10^{-5}$	$7.7893 \cdot 10^{-7}$	$7.8214 \cdot 10^{-7}$
estimateur avec recyclage (2.3.1.2)	$1.28455 \cdot 10^{-5}$	$3.1464 \cdot 10^{-6}$	$3.11887 \cdot 10^{-6}$
estimateur mixte (§ 2.4.3.4)	$1.159 \cdot 10^{-5}$		$1.0491 \cdot 9.2^{-7}$

TAB. 4.4.7 – GT-MHR - Probabilité de rupture

nous ne pouvons en déduire la comparaison entre les deux concepts, les conditions n'étant pas encore représentative et le taux de combustion final moyen sensiblement différent (8.33% FIMA pour PBMR et 9.56% FIMA pour GT-MHR).

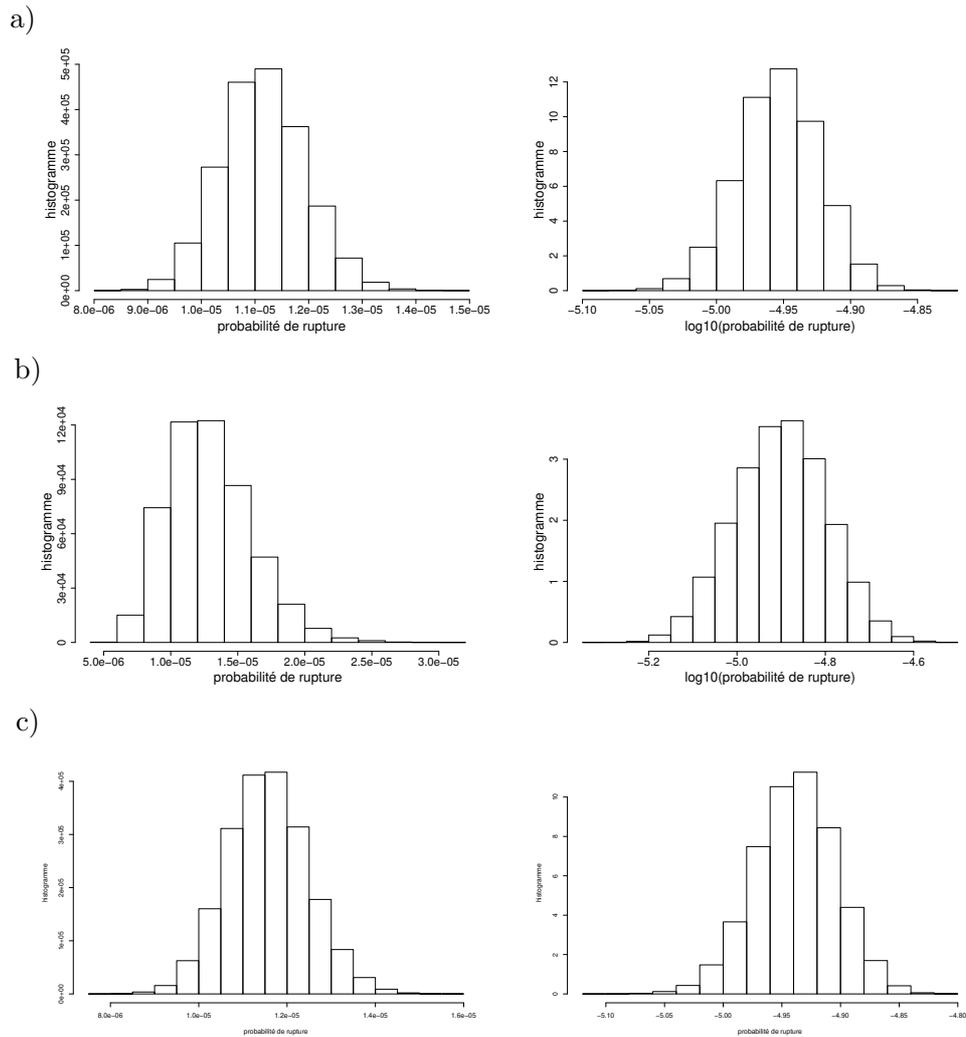


FIG. 4.4.10 – Histogramme de la probabilité de rupture estimée obtenu par bootstrap sur 50000 ré-échantillonnages - GT-MHR. a) estimateur sans recyclage (2.3.1.1), b) estimateur avec recyclage (2.3.1.2), c) estimateur mixte (§ 2.4.3.4)

## CHAPITRE 5

### ESTIMATION DE QUANTILES ET APPLICATION

#### 5.1 Introduction

L'estimation de quantiles est un des problèmes fondamentaux en statistique, et représente un enjeu important pour les applications [LWDK 91]. Dans des situations complexes, par exemple dans le cas où l'on recherche le quantile d'une quantité qui est une sortie d'un gros code numérique, le nombre d'appels au code peut être limité et rendre difficile l'application de méthodes de type Monte Carlo. Des techniques de réduction de variance ont été mises au point et implémentées pour répondre à ce genre de problèmes. Ces techniques sont en général basées sur des méthodes d'échantillonnage préférentiel [Gly 96], des méthodes de corrélation-induction [AW 88], et des méthodes de variables de contrôle [HeN 98]. Dans ce chapitre, nous allons nous concentrer sur les techniques de réduction de variance utilisant un modèle réduit. La configuration que nous avons en tête est celle où nous disposons, en plus du gros code numérique, d'un code simplifié qui fournit une réponse approchée, mais qui est beaucoup moins chère en termes de coût de calcul. Nous allons voir comment exploiter au mieux l'existence d'un tel modèle réduit.

L'estimation du quantile d'une variable aléatoire réelle  $Y$  a pour but de déterminer le niveau  $y \in \mathbb{R}$  telle que la probabilité que  $Y$  soit plus petite que  $y$  ait une valeur prescrite. Nous supposons dans ce chapitre que  $Y$  est une v.a. à densité et que la densité de  $Y$  est strictement positive sur  $\mathbb{R}$  (ou sur l'intervalle  $I$  supportant la densité). On recherche une estimation du  $\alpha$ -quantile de  $Y$ , c'est-à-dire du réel  $y_\alpha$  tel que  $\mathbb{P}(Y \leq y_\alpha) = \alpha$ .

Dans ce chapitre, on considérera plus particulièrement le cas où  $Y$  est la (une) sortie réelle d'un gros code numérique  $f$ , dont les paramètres d'entrée sont aléatoires et modélisés par le vecteur aléatoire  $X \in \mathbb{R}^d$ , dont la loi est connue. Notre motivation vient des progrès récents en calcul scientifique et en méthodes numériques, qui ont permis la mise au point de codes de calculs numériques modélisant des phénomènes de plus en plus complexes [SMW 89, FLS 06]. Le problème de l'estimation de quantiles peut être résolu par des techniques d'échantillonnage classique de type Monte Carlo, ou par des méthodes de quasi Monte Carlo (du type hypercube latin ou autres). Lorsque les simulations numériques sont trop coûteuses pour obtenir une estimation du quantile recherché avec une précision suffisante, des méthodes basées sur des

résultats de statistique d'ordre sont possibles [Dav 81, NeWGB 04]. Une alternative consiste à remplacer le gros code numérique par un modèle réduit, du type métamodèle ou surface de réponse [LWDK 91, FLS 06].

Cependant, il faut prendre garde que l'estimation d'un quantile à partir d'un métamodèle ou d'une surface de réponse peut conduire à des erreurs importantes, car ces modèles réduits sont en général construits par lissage de résultats issus du gros code numérique. Récemment, certains auteurs ont exploité les avantages d'un type particulier de métamodèle, à base de processus gaussien [SMW 89], qui donne non seulement un prédicteur (la moyenne) d'une simulation numérique, mais aussi un indicateur local sur la précision de la prédiction (la variance). Dans ce cadre, les auteurs de [Oak 04] et [VPM ] ont développé des procédures séquentielles pour choisir des points de conception et pour construire des métamodèles gaussiens précis près des zones d'importance, dont dépend la valeur du quantile. Rutherford [Rut 06] a aussi proposé d'utiliser des techniques de simulation conditionnelle, utilisées surtout en géostatistique, pour obtenir plusieurs réalisations possibles du processus gaussien, et pour en déduire un estimateur du quantile non-biaisé. Toutes ces techniques sont basées sur la construction d'un modèle à base de processus gaussien, ce qui peut s'avérer difficile en grande dimension ( $d > 10$ ) ou pour des phénomènes discontinus. De plus, il apparaît en pratique très souvent qu'un métamodèle est souvent disponible dans les applications industrielles, ce modèle provenant d'une version antérieure du code principal ou d'un modèle physique simplifié.

Dans ce chapitre, nous supposons donc que nous disposons d'un code complet  $f$ , dont chaque appel est coûteux, et d'un modèle réduit  $f_r$  beaucoup plus léger mais approximatif. Nous ne chercherons pas à construire un métamodèle plus élaboré, mais nous chercherons à utiliser au mieux le modèle réduit  $Z = f_r(X)$  pour mettre au point une méthode de réduction de variance pour l'estimation d'un quantile de  $Y = f(X)$ .

## 5.2 Estimation de quantile sur un $n$ -échantillon

On dispose d'un  $n$ -échantillon  $(Y_1, \dots, Y_n)$  de variables aléatoires indépendantes identiquement distribuées (v.a. i.i.d.) selon une loi inconnue à densité  $p(y)$ . On cherche un estimateur du  $\alpha$ -quantile  $y_\alpha$  défini par :

$$\mathbb{P}(Y \leq y_\alpha) = \alpha$$

Un estimateur est bâti sur une fonction  $\Theta_n : \mathbb{R}^n \rightarrow \mathbb{R}$  à définir, qui doit être telle que  $\hat{Y}_{\alpha,n} := \Theta_n(Y_1, \dots, Y_n)$  est "proche" de la valeur  $y_\alpha$ .

Pour quantifier cette notion de proximité, on utilise classiquement la notion de risque

quadratique moyen, somme du carré du biais :

$$\mathbb{E}[\hat{Y}_{\alpha,n}] - y_\alpha$$

et de la variance :

$$\text{Var}(\hat{Y}_{\alpha,n})$$

On peut réclamer que l'estimateur soit sûr au niveau  $\beta$  au sens où

$$\mathbb{P}(\hat{Y}_{\alpha,n} \geq y_\alpha) \geq \beta \quad (5.2.1)$$

Bien sûr, si  $\hat{Y}_{\alpha,n}$  est un estimateur sûr au niveau  $\beta$  au sens (5.2.1), alors tout autre estimateur  $\hat{Y}'_{\alpha,n}$  tel que  $\hat{Y}'_{\alpha,n} \geq \hat{Y}_{\alpha,n}$  le sera aussi. On cherche donc, parmi tous les estimateurs sûrs au niveau  $\beta$ , ceux qui sont les plus petits.

Les estimateurs qui suivent sont basés sur des notions de statistique d'ordre. En clair, on associe à l'échantillon  $(Y_1, \dots, Y_n)$  l'échantillon ordonné  $(Y_{(1)}, \dots, Y_{(n)})$  tel que  $Y_{(1)} < \dots < Y_{(n)}$ .

### 5.2.1 Quantile empirique

L'estimateur classique du  $\alpha$ -quantile est le quantile empirique

$$\hat{Y}_{\alpha,n} = Y_{([\alpha n]+1)} \quad (5.2.2)$$

où  $[x]$  est la partie entière de  $x$ . La version la plus raffinée de ce résultat est basée sur une formule d'interpolation qu'on ne développera pas ici. On supposera dans la suite que  $\alpha n$  est un entier.

Le résultat asymptotique ( $n \rightarrow \infty$ ) est le suivant (dans le cas où la densité  $p(y)$  est dérivable en  $y_\alpha$ ), donné par exemple dans [Dav 81].

$\hat{Y}_{\alpha,n}$  est un estimateur biaisé et asymptotiquement normal :

$$\mathbb{E}[\hat{Y}_{\alpha,n}] = y_\alpha - \frac{\alpha(1-\alpha)p'(y_\alpha)}{2(n+2)p^3(y_\alpha)} + O\left(\frac{1}{n^2}\right) \quad (5.2.3)$$

$$\text{Var}(\hat{Y}_{\alpha,n}) = \frac{\alpha(1-\alpha)}{(n+2)p^2(y_\alpha)} + O\left(\frac{1}{n^2}\right) \quad (5.2.4)$$

Le biais est donc asymptotiquement négligeable (comparé à l'écart-type), et la variance est d'autant plus grande que l'on cherche à évaluer un quantile extrême (la densité au point  $y_\alpha$  est alors petite).

Construire un estimateur sûr au niveau  $\beta$  en utilisant le résultat précédent n'est pas une bonne idée. En effet, on pourrait proposer

$$\hat{Y}_{\alpha,n} + t_\beta \frac{\sqrt{\alpha(1-\alpha)}}{\sqrt{n+2}p(y_\alpha)}$$

avec  $t_\beta = \Phi^{-1}(\beta)$  et  $\Phi$  la fonction de répartition de la loi normale centrée réduite. Cependant, on voit qu'il faudrait alors estimer  $p(y_\alpha)$ . L'estimation d'une densité, surtout dans la queue d'une distribution, n'est pas commode.

### 5.2.2 Quantile de Wilks

Le quantile empirique (5.2.2) vérifiant

$$\sqrt{n}(\hat{Y}_{\alpha,n} - y_\alpha) \xrightarrow{n \rightarrow \infty} \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{p^2(y_\alpha)}\right)$$

on peut en déduire que, asymptotiquement, il vérifie  $\mathbb{P}(\hat{Y}_{\alpha,n} \geq y_\alpha) \simeq 1/2$ . Pour augmenter cette probabilité, on peut penser à prendre une certaine marge et à proposer un estimateur de la forme

$$\hat{Y}_{\alpha,n} = Y_{(n\alpha+r)}$$

La question est de savoir quelle marge il faut prendre, c'est-à-dire quel entier  $r$  il faut choisir pour que  $\mathbb{P}(Y_{(n\alpha+r)} \geq y_\alpha) \geq \beta$ .

Il est possible de choisir le  $r$  de manière optimale, grâce au résultat suivant qui tire sa puissance de sa grande simplicité. Il stipule que le nombre de dépassements d'un seuil  $y$  par la suite de v.a. i.i.d.  $(Y_1, \dots, Y_n)$  suit une loi binomiale de paramètres  $(n, q)$ , avec  $q = \mathbb{P}(Y > y)$ .

*Quels que soient  $0 \leq j \leq n$ ,*

$$\mathbb{P}(j \text{ des } Y_i \text{ sont } > y) = C_n^j (1 - F(y))^j F(y)^{n-j}$$

où  $F(y) = \mathbb{P}(Y \leq y)$  est la fonction de répartition des  $Y_i$ .

En prenant  $y = y_\alpha$  on obtient :

$$\mathbb{P}(j \text{ des } Y_i \text{ sont } > y_\alpha) = C_n^j (1 - \alpha)^j \alpha^{n-j}$$

Or l'événement  $\{Y_{(k)} > y_\alpha\}$  se produit si et seulement si au moins  $n - k + 1$  des  $Y_i$  sont  $> y_\alpha$ .

Par conséquent

$$\begin{aligned} \mathbb{P}(Y_{(n\alpha+r)} > y_\alpha) &= \sum_{j=n(1-\alpha)-r+1}^n \mathbb{P}(j \text{ des } Y_i \text{ sont } > y_\alpha) \\ &= \sum_{j=n(1-\alpha)-r+1}^n C_n^j (1 - \alpha)^j \alpha^{n-j} \\ &= 1 - C_\alpha(n, r) \end{aligned}$$

avec

$$C_\alpha(n, r) = \sum_{j=0}^{n(1-\alpha)-r} C_n^j (1 - \alpha)^j \alpha^{n-j} \quad (5.2.5)$$

Conclusion : Si on note  $r$  le plus petit entier tel que

$$C_\alpha(n, r) \leq 1 - \beta \quad (5.2.6)$$

alors  $\mathbb{P}(Y_{(\alpha n+r)} > y_\alpha) \geq \beta$ , c'est-à-dire que l'estimateur  $Y_{(\alpha n+r)}$  est sûr au niveau  $\beta$ .

En fait, en suivant les parties entières, on voit qu'il est préférable d'utiliser la formule d'interpolation

$$\hat{Y}_{\alpha, n} = a_1 Y_{(\alpha n+r)} + a_2 Y_{(\alpha n+r-1)} \quad (5.2.7)$$

avec

$$a_2 = \frac{1 - \beta - C_\alpha(n, r)}{C_\alpha(n, r-1) - C_\alpha(n, r)}$$

$$a_1 = -\frac{1 - \beta - C_\alpha(n, r-1)}{C_\alpha(n, r-1) - C_\alpha(n, r)}$$

Application : Si  $n = 200$ ,  $\alpha = \beta = 0.95$ , alors  $\alpha n + r = 196$ ,  $a_1 = 0.34$ ,  $a_2 = 0.66$ .

L'estimateur de Wilks est valable même pour  $n$  petit, tout au moins pour  $n \geq n_c$  où  $n_c$  est la taille critique de l'échantillon à partir duquel l'estimateur de Wilks est bien défini. L'existence d'un tel  $n$  critique est naturel : en effet on ne peut pas être certain que le maximum absolu d'un échantillon de taille 10 soit au dessus du quantile à 99%. La taille critique  $n_c$  est le plus petit  $n$  tel que  $\mathbb{P}(Y_{(n)} > y_\alpha) \geq \beta$ , c'est-à-dire

$$n_c = \left\lceil \frac{\ln(1 - \beta)}{\ln(\alpha)} \right\rceil + 1$$

Application : Si  $\alpha = \beta = 0.95$ , alors  $n_c = 59$ .

On peut décrire le comportement asymptotique de l'indice  $r$  de l'estimateur de Wilks lorsque  $n \rightarrow \infty$ . On trouve que

$$r \sim t_\beta \sqrt{\alpha(1 - \alpha)n} \quad (5.2.8)$$

avec  $t_\beta = \Phi^{-1}(\beta)$ .

Application : pour  $\beta = 0.95$ , on a  $t_\beta \simeq 1.645$ . Si de plus  $\alpha = 0.95$  et  $n = 200$ , on trouve  $t_\beta \sqrt{\alpha(1 - \alpha)n} \simeq 5.07$ , ce qui correspond à la vraie valeur 6 (en prenant l'entier supérieur).

*Démonstration* : On donne ci-dessous la preuve de (5.2.8).

On prend  $r = r' \sqrt{n}$  et on écrit

$$C_\alpha(n, r) = \sum_{j=0}^{n(1-\alpha)-r'\sqrt{n}} C_n^j (1 - \alpha)^j \alpha^{n-j}$$

On va calculer l'asymptotique de cette expression quand  $n \rightarrow \infty$ . On utilise pour cela la formule de Stirling, et on change d'indice  $k = j - n(1 - \alpha)$  :

$$C_\alpha(n, r) \simeq \frac{1}{\sqrt{2\pi n \alpha(1 - \alpha)}} \sum_{k=-n(1-\alpha)}^{-r'\sqrt{n}} \left(1 + \frac{k}{n(1 - \alpha)}\right)^{-n(1-\alpha)-k} \left(1 - \frac{k}{n\alpha}\right)^{-n\alpha+k}$$

On prend ensuite la forme exponentielle, et on développe le logarithme à l'ordre deux :

$$\begin{aligned}
C_\alpha(n, r) &\simeq \frac{1}{\sqrt{2\pi n\alpha(1-\alpha)}} \sum_{k=-n(1-\alpha)}^{-r'\sqrt{n}} \exp \left[ -(n(1-\alpha) + k) \log \left( 1 + \frac{k}{n(1-\alpha)} \right) \right. \\
&\quad \left. -(n\alpha - k) \log \left( 1 - \frac{k}{n\alpha} \right) \right] \\
&\simeq \frac{1}{\sqrt{2\pi n\alpha(1-\alpha)}} \sum_{k=-n(1-\alpha)}^{-r'\sqrt{n}} \exp \left[ -(n(1-\alpha) + k) \left( \frac{k}{n(1-\alpha)} - \frac{k^2}{2n^2(1-\alpha)^2} \right) \right. \\
&\quad \left. -(n\alpha - k) \left( -\frac{k}{n\alpha} - \frac{k^2}{2n^2\alpha^2} \right) \right] \\
&\simeq \frac{1}{\sqrt{2\pi n\alpha(1-\alpha)}} \sum_{k=-n(1-\alpha)}^{-r'\sqrt{n}} \exp \left[ -\frac{k^2}{2n} \left( \frac{1}{\alpha} + \frac{1}{1-\alpha} \right) \right]
\end{aligned}$$

On reconnaît une somme de Riemann

$$C_\alpha(n, r) \simeq \frac{1}{\sqrt{2\pi\alpha(1-\alpha)}} \int_{-\infty}^{-r'} \exp \left[ -\frac{s^2}{2\alpha(1-\alpha)} \right] ds$$

ce qui donne

$$C_\alpha(n, r) \simeq \Phi \left( \frac{-r'}{\sqrt{\alpha(1-\alpha)}} \right) = 1 - \Phi \left( \frac{r'}{\sqrt{\alpha(1-\alpha)}} \right)$$

■

Concluons ce paragraphe en fournissant quelques comparaisons entre estimateur empirique et estimateur de Wilks pour différentes lois (figure 5.2.1).

### 5.2.3 Quantile empirique avec bootstrap

La méthode bootstrap consiste à ré-échantillonner dans l'échantillon lui-même pour examiner les fluctuations des estimateurs empiriques [Efr 79].

A partir de l'échantillon  $(Y_1, \dots, Y_n)$ , on tire une série très grande d'échantillons de même taille  $(\tilde{Y}_1^l, \dots, \tilde{Y}_n^l)$ ,  $l = 1, \dots, M$ ,  $M \gg 1$ , de la manière suivante. Conditionnellement à  $(Y_1, \dots, Y_n)$ , les  $\tilde{Y}_i^l$ ,  $l = 1, \dots, M$ ,  $i = 1, \dots, n$  sont i.i.d. de loi uniforme sur l'ensemble des  $n$  points  $(Y_1, \dots, Y_n)$  :

$$\mathbb{P} \left( \tilde{Y} = Y_1 | (Y_1, \dots, Y_n) \right) = \dots = \mathbb{P} \left( \tilde{Y} = Y_n | (Y_1, \dots, Y_n) \right) = \frac{1}{n}$$

Pour chaque  $l = 1, \dots, M$ , on considère l'échantillon ordonné  $(\tilde{Y}_{(1)}^l, \dots, \tilde{Y}_{(n)}^l)$ . Les v.a.  $\tilde{Y}_{(\alpha n)}^l$  sont identiquement distribuées. Le résultat asymptotique ( $n \rightarrow \infty$ ) est le suivant [BF 81, Fre 81].

$\sqrt{n}(\tilde{Y}_{(\alpha n)} - y_\alpha)$  a la même distribution que  $\sqrt{n}(Y_{(\alpha n)} - y_\alpha)$  quand  $n \rightarrow \infty$ .

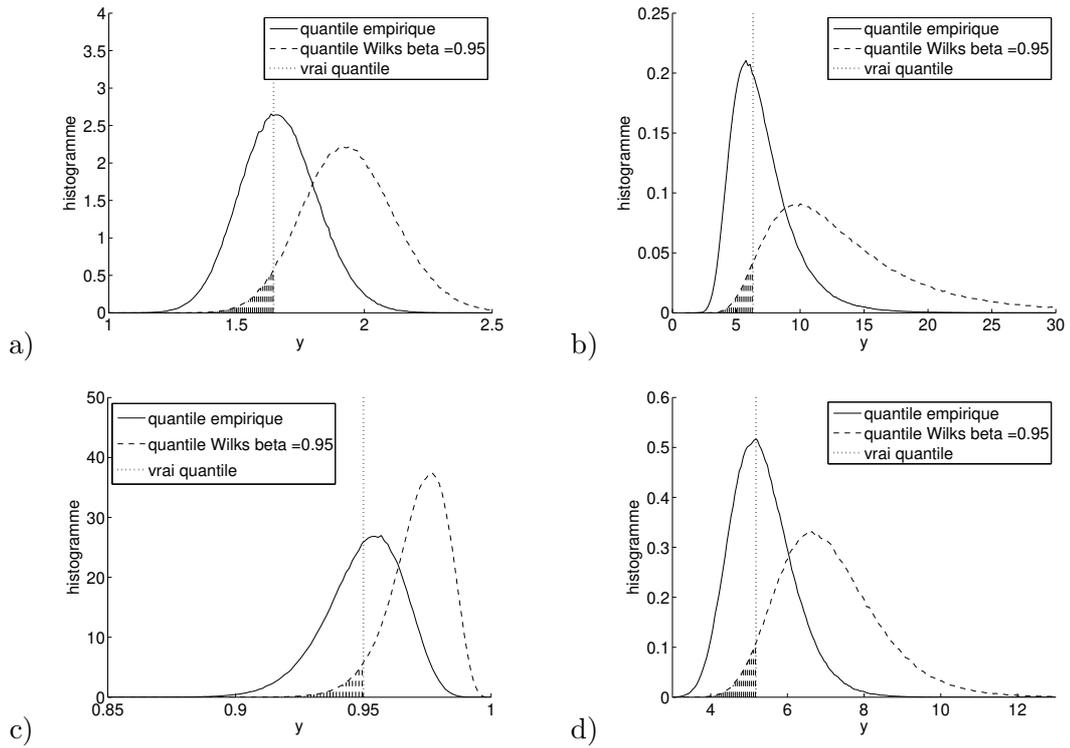


FIG. 5.2.1 – Comparaisons entre l'estimateur empirique et l'estimateur de Wilks au niveau  $\beta$  pour le  $\alpha$ -quantile d'une v.a.  $Y$ . Ici  $\alpha = \beta = 0.95$  et l'échantillon a pour taille  $n = 200$ . Les histogrammes des estimateurs sont tracés à partir de  $10^6$  expériences. La loi de  $Y$  est une gaussienne  $\mathcal{N}(0, 1)$  (figure a), une Cauchy (figure b), une loi uniforme sur  $[0, 1]$  (figure c) et une log-normale (figure d). Comme prédit par la théorie, l'estimateur de Wilks donne une valeur inférieure au vrai quantile dans 5% des simulations (zones hachurées). Cette précision a un prix : la dispersion de l'estimateur de Wilks est supérieure à celle de l'estimateur empirique.

La distribution conditionnelle du  $\alpha$ -quantile empirique avec bootstrap  $\tilde{Y}_{(n\alpha)}$ , conditionnellement à  $(Y_1, \dots, Y_n)$ , se calcule de manière similaire à la formule de Wilks. On a en effet

$$\mathbb{P}(j \text{ des } \tilde{Y}_i \text{ sont } > y | (Y_1, \dots, Y_n)) = C_n^j (1 - F_n(y))^j F_n(y)^{n-j}$$

où  $F_n(y)$  est la fonction de répartition de  $\tilde{Y}$  conditionnellement à  $(Y_1, \dots, Y_n)$  :

$$F_n(y) = \mathbb{P}(\tilde{Y} \leq y | (Y_1, \dots, Y_n)) = \frac{k}{n} \text{ si } y \in [Y_{(k)}, Y_{(k+1)}[$$

avec la convention  $Y_{(0)} = -\infty$  et  $Y_{(n+1)} = +\infty$ . Par conséquent, la loi de  $\tilde{Y}_{(n\alpha)}$ , conditionnel-

lement à  $(Y_1, \dots, Y_n)$ , est donnée par :

$$\begin{aligned} \mathbb{P}(\tilde{Y}_{(n\alpha)} > y | (Y_1, \dots, Y_n)) &= \sum_{j=n(1-\alpha)+1}^n \mathbb{P}(j \text{ des } \tilde{Y}_i \text{ sont } > y | (Y_1, \dots, Y_n)) \\ &= 1 - \sum_{j=0}^{n(1-\alpha)} C_n^j (1 - F_n(y))^j F_n(y)^{n-j} \end{aligned}$$

et donc

$$\mathbb{P}(\tilde{Y}_{(n\alpha)} > y | (Y_1, \dots, Y_n)) = 1 - \sum_{j=0}^{n(1-\alpha)} C_n^j \left(1 - \frac{k}{n}\right)^j \left(\frac{k}{n}\right)^{n-j} \quad (5.2.9)$$

pour tout  $y \in [Y_{(k)}, Y_{(k+1)}[$ .

On observe le  $\hat{y}_{\alpha,n}$  tel que une fraction  $1 - \beta$  des échantillons bootstrap  $\tilde{Y}_{(n\alpha)}^l$ ,  $l = 1, \dots, M$ , sont au delà de  $\hat{y}_{\alpha,n}$ . Le résultat sur le comportement asymptotique des estimateurs bootstrap énoncé ci-dessus nous dit que  $\hat{y}_{\alpha,n}$  est un estimateur de  $y_\alpha$  sûr au niveau  $\beta$ , dans l'asymptotique  $n \rightarrow \infty$ . En fait, on peut déterminer cet estimateur de manière précise ici. En effet, supposons que  $M$  soit suffisamment grand pour qu'on puisse considérer qu'on estime parfaitement la loi de  $\tilde{Y}_{(n\alpha)}$  conditionnellement à  $(Y_1, \dots, Y_n)$ . Alors, d'après (5.2.9), on a simplement

$$\hat{y}_{\alpha,n} = Y_{(n-k)}$$

avec  $k$  le plus grand entier tel que

$$\sum_{j=0}^{n(1-\alpha)} C_n^j \left(\frac{k}{n}\right)^j \left(1 - \frac{k}{n}\right)^{n-j} \geq \beta \quad (5.2.10)$$

Le  $\hat{y}_{\alpha,n}$  ainsi obtenu est très proche de l'estimateur de Wilks. En fait, dans l'asymptotique  $n \rightarrow \infty$ , on trouve que

$$n - k \simeq \alpha n + t_\beta \sqrt{\alpha(1-\alpha)n}$$

Le preuve est très similaire à celle de (5.2.8) : On calcule l'asymptotique de (5.2.10) avec  $k = n(1-\alpha) + r'\sqrt{n}$  avec la formule de Stirling, on reconnaît une somme de Riemann et on la calcule par son approximation sous forme d'intégrale continue.

L'indice du quantile empirique avec bootstrap a exactement le même comportement asymptotique que l'indice de l'estimateur de Wilks obtenu dans la section précédente. Donc le quantile empirique avec bootstrap est asymptotiquement équivalent au quantile de Wilks quand  $n \rightarrow \infty$ . Pour  $n$  fini, l'estimateur de Wilks est cependant à préférer car on est toujours assuré qu'il répond à la spécification d'être sûr au niveau  $\beta$ . De plus, il est beaucoup plus simple à mettre en œuvre que le quantile bootstrap.

### 5.2.4 Autres résultats sur les dépassements de seuil

On suppose ici comme précédemment que l'on dispose d'un échantillon  $(Y_1, \dots, Y_n)$  de taille  $n$ . On note encore  $(Y_{(1)}, \dots, Y_{(n)})$  la statistique d'ordre associée.

#### 5.2.4.1 Prédiction de dépassement d'un seuil

On reformule la question de l'estimation de dépassement d'un seuil sous la forme suivante. Quelle est la probabilité pour que, parmi  $r$  nouvelles simulations, il y en ait  $j$  qui dépassent le niveau  $Y_{(k)}$ ? On note  $N_{n,k}^r$  la variable aléatoire qui décrit le nombre de simulations parmi les  $r$  nouvelles qui dépassent le niveau  $Y_{(k)}$ . Cette variable est à valeurs dans  $\{0, \dots, r\}$  et suit une loi hypergéométrique :

$$\mathbb{P}(N_{n,k}^r = j) = \frac{C_{n-k+r-j}^{n-k} C_{k-1+j}^{k-1}}{C_{r+n}^n}, \quad j = 0, \dots, r$$

En particulier, la probabilité

$$\mathbb{P}(N_{n,k}^r = 0) = \frac{n(n-1)\cdots(n-k+1)}{(r+n)(r+n-1)\cdots(r+n-k+1)}$$

est la probabilité que le maximum obtenu lors de  $r$  simulations soit inférieur à  $Y_{(k)}$  :

$$\mathbb{P}(N_{n,k}^r = 0) = \mathbb{E}[(1 - F(Y_{(k)}))^r]$$

Cette formule permet de calculer les moments de  $F(Y_{(k)})$  pour tout  $1 \leq k \leq n$ . On a ainsi, pour les deux premiers moments :

$$\mathbb{E}[F(Y_{(k)})] = \frac{k}{n+1} \tag{5.2.11}$$

$$\text{Var}(F(Y_{(k)})) = \frac{k(n-k+1)}{(n+1)^2(n+2)} \tag{5.2.12}$$

Les fluctuations de  $Y_{(k)}$  observées dans l'échelle  $F$  ont donc une forme universelle, contrairement à ces mêmes fluctuations observées dans leur échelle naturelle.

#### 5.2.4.2 Temps d'attente d'un dépassement

Fixons un seuil  $y \in \mathbb{R}$  et notons  $i_1 < i_2 < \dots < i_k$  les indices pour lesquels  $Y_i > y$ . La suite des incréments  $\Delta_j = i_j - i_{j-1}$ ,  $j = 1, \dots, k$  (avec la convention  $i_0 = 0$ ) est une suite de v.a. i.i.d. de loi géométrique de paramètre  $1 - F(y)$  :

$$\mathbb{P}(\Delta_j = i) = F(y)^{i-1}(1 - F(y)), \quad i \geq 1$$

L'observation de cette suite permet d'estimer  $F(y)$ .

Ce résultat peut servir pour vérifier le caractère i.i.d. d'une suite de v.a. Il suffit d'utiliser un test classique d'adéquation de loi pour tester si la suite des incréments est une suite de v.a. géométriques.

### 5.3 Estimation de quantile par variable de contrôle

On cherche à évaluer le  $\alpha$ -quantile de  $Y$ . Par rapport à la section précédente, on suppose ici que  $Y = f(X)$  est la variable de sortie d'un gros code numérique prenant en entrée un vecteur aléatoire  $X$ , et qu'on dispose d'un modèle réduit (appelé aussi surface de réponse ou métamodèle)  $Z = f_r(X)$  qu'on connaît bien. On suppose ici que  $f$  et  $f_r$  sont toutes deux à valeurs dans  $\mathbb{R}$ . On est capable de faire un grand nombre de calculs avec le modèle réduit, et on connaît les  $\alpha$ -quantiles  $z_\alpha$  de  $Z$  pour tout  $\alpha$ . On va utiliser la v.a.  $Z$  comme contrôle pour réduire la variance de l'estimateur du quantile de  $Y$ .

On va proposer deux estimateurs de la fonction de répartition, l'un basé sur un principe de maximum de vraisemblance, l'autre sur la méthode usuelle de réduction de variance pour des simulations de Monte Carlo avec variable de contrôle. On va voir que ces deux estimateurs sont en fait identiques, on va évaluer leurs propriétés asymptotiques, puis on va les utiliser pour estimer le  $\alpha$ -quantile de  $Y$ .

#### 5.3.1 Estimation de la fonction de répartition par maximum de vraisemblance

On reprend dans cette section les idées de Hsu and Nelson [HN 87]. On note, pour tout  $y \in \mathbb{R}$ ,

$$p_{00}(y) = \mathbb{P}(Z \leq z_\alpha, Y \leq y) \quad p_{01}(y) = \mathbb{P}(Z \leq z_\alpha, Y > y)$$

$$p_{10}(y) = \mathbb{P}(Z > z_\alpha, Y \leq y) \quad p_{11}(y) = \mathbb{P}(Z > z_\alpha, Y > y)$$

Ces quatre probabilités sont liées par les relations

$$p_{00}(y) + p_{01}(y) = \mathbb{P}(Z \leq z_\alpha) = \alpha \tag{5.3.1}$$

$$p_{10}(y) + p_{11}(y) = \mathbb{P}(Z > z_\alpha) = 1 - \alpha \tag{5.3.2}$$

Si on arrive à estimer les  $p_{jl}(y)$ , alors on aura un estimateur de  $F(y) = \mathbb{P}(Y \leq y)$  car

$$F(y) = p_{00}(y) + p_{10}(y)$$

Considérons un échantillon  $((Y_1, Z_1), \dots, (Y_n, Z_n))$  de v.a. i.i.d. de même loi que  $(Y, Z)$ . Fixons  $y \in \mathbb{R}$  et notons

$$N_{00}(y) = \text{Card} \{i = 1, \dots, n, Z_i \leq z_\alpha, Y_i \leq y\}$$

$$N_{01}(y) = \text{Card} \{i = 1, \dots, n, Z_i \leq z_\alpha, Y_i > y\}$$

$$N_{10}(y) = \text{Card} \{i = 1, \dots, n, Z_i > z_\alpha, Y_i \leq y\}$$

$$N_{11}(y) = \text{Card} \{i = 1, \dots, n, Z_i > z_\alpha, Y_i > y\}$$

Les v.a.  $N_0 = \text{Card} \{i = 1, \dots, n, Z_i \leq z_\alpha\}$  et  $N_1 = \text{Card} \{i = 1, \dots, n, Z_i > z_\alpha\}$  vérifient  $N_0 + N_1 = n$  et

$$N_{00} + N_{01} = N_0, \quad N_{10} + N_{11} = N_1$$

La loi jointe de  $(N_{00}, N_{01}, N_{10}, N_{11})$  est multinomiale :

$$\mathbb{P}(N_{jl} = n_{jl}, j, l = 0, 1) = \frac{n!}{\prod_{j,l=0}^1 n_{jl}!} \prod_{j,l=0}^1 [p_{jl}(y)]^{n_{jl}}$$

En tenant compte du fait que les  $p_{jl}(y)$  sont liés par les relations (5.3.1-5.3.2), le membre de droite ne fait apparaître que les deux paramètres  $p_{00}(y)$  et  $p_{10}(y)$  :

$$\mathbb{P}(N_{jl} = n_{jl}, j, l = 0, 1) = \frac{n!}{\prod_{j,l=0}^1 n_{jl}!} [p_{00}(y)]^{n_{00}} [\alpha - p_{00}(y)]^{n_{01}} [p_{10}(y)]^{n_{10}} [1 - \alpha - p_{10}(y)]^{n_{11}}$$

On trouve alors facilement que les estimateurs du maximum de vraisemblance pour  $p_{00}(y)$  et  $p_{10}(y)$  sont :

$$\hat{p}_{00}(y) = \frac{\alpha N_{00}(y)}{N_0}, \quad \hat{p}_{10}(y) = \frac{(1 - \alpha) N_{10}(y)}{N_1}, \quad (5.3.3)$$

conditionnellement en  $N_0 > 0$  et  $N_1 > 0$ . On en déduit les estimateurs de  $p_{01}(y)$  et  $p_{11}(y)$  :

$$\hat{p}_{01}(y) = \frac{\alpha N_{01}(y)}{N_0}, \quad \hat{p}_{11}(y) = \frac{(1 - \alpha) N_{11}(y)}{N_1}. \quad (5.3.4)$$

Un estimateur de  $F(y)$  est donc :

$$\hat{F}(y) = \hat{p}_{00}(y) + \hat{p}_{10}(y) = \frac{\alpha N_{00}(y)}{N_0} + \frac{(1 - \alpha) N_{10}(y)}{N_1} \quad (5.3.5)$$

### 5.3.2 Estimation de la fonction de répartition avec contrôle

Un estimateur par variable de contrôle  $Z$  de  $F(y)$  est de la forme

$$\hat{F}(y) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{Y_j \leq y} - \hat{\beta} \left( \frac{1}{n} \sum_{j=1}^n g(Z_j) - \mathbb{E}[g(Z)] \right) \quad (5.3.6)$$

où la fonction  $g$  est à choisir par l'utilisateur [Nel 80]. L'espérance  $\mathbb{E}[g(Z)]$  est supposée connue. Le paramètre  $\beta$  optimal est le coefficient de corrélation de  $g(Z)$  et  $\mathbf{1}_{Y \leq y}$ , qui est inconnu. On utilise alors le paramètre  $\hat{\beta}$  défini comme la pente obtenue par régression linéaire des  $\mathbf{1}_{Y_j \leq y}$  sur les  $g(Z_i)$  (méthode des moindres carrés) :

$$\hat{\beta} = \frac{\sum_{j=1}^n (\mathbf{1}_{Y_j \leq y} - \hat{F}_n(y))(g(Z_j) - \hat{g}_n)}{\sum_{j=1}^n (g(Z_j) - \hat{g}_n)^2}$$

avec

$$\hat{F}_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{Y_i \leq y}, \quad \hat{g}_n = \frac{1}{n} \sum_{i=1}^n g(Z_i)$$

On trouve alors que l'estimateur  $\hat{F}(y)$  peut se réécrire comme une somme pondérée

$$\hat{F}(y) = \sum_{j=1}^n W_j \mathbf{1}_{Y_j \leq y}$$

avec

$$W_j = \frac{1}{n} + \frac{(\hat{g}_n - \mathbb{E}[g(Z)])(\hat{g}_n - g(Z_j))}{\sum_{i=1}^n (g(Z_i) - \hat{g}_n)^2}, \quad \hat{g}_n = \frac{1}{n} \sum_{i=1}^n g(Z_i)$$

Notez que  $\sum_{j=1}^n W_j = 1$ .

Si on choisit  $g(z) = \mathbf{1}_{z \leq z_\alpha}$ , alors  $\mathbb{E}[g(Z)] = \alpha$ ,  $\hat{g}_n = N_0/n$  (avec les notations de la section précédente) et on trouve

$$W_j = \frac{\alpha}{N_0} \mathbf{1}_{Z_j \leq z_\alpha} + \frac{1-\alpha}{N_1} \mathbf{1}_{Z_j > z_\alpha} \quad (5.3.7)$$

L'estimateur de la fonction de répartition de  $Y$  est

$$\hat{F}(y) = \sum_{j=1}^n W_j \mathbf{1}_{Y_j \leq y} = N_{00}(y) \frac{\alpha}{N_0} + N_{10}(y) \frac{1-\alpha}{N_1} \quad (5.3.8)$$

On retrouve exactement l'estimateur (5.3.5) proposé dans la section précédente.

En utilisant les résultats classiques sur la réduction de variance pour les simulations de Monte Carlo [Nel 80], on trouve que l'estimateur avec variable de contrôle (VC)

$$\sqrt{n}(\hat{F}(y) - F(y)) |_{\text{VC}} \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, \sigma^2), \quad \sigma^2 = F(y)(1-F(y))(1-\rho_I^2) \quad (5.3.9)$$

où  $\rho_I$  est le coefficient de corrélation entre  $\mathbf{1}_{Y \leq y}$  et  $\mathbf{1}_{Z \leq z_\alpha}$  :

$$\begin{aligned} \rho_I &= \frac{\mathbb{P}(Y \leq y, Z \leq z_\alpha) - \mathbb{P}(Y \leq y)\mathbb{P}(Z \leq z_\alpha)}{\sqrt{(1-\mathbb{P}(Y \leq y))\mathbb{P}(Y \leq y)(1-\mathbb{P}(Z \leq z_\alpha))\mathbb{P}(Z \leq z_\alpha)}} \\ &= \frac{\mathbb{P}(Y \leq y, Z \leq z_\alpha) - \alpha F(y)}{\sqrt{F(y)(1-F(y))}\sqrt{\alpha - \alpha^2}} \end{aligned} \quad (5.3.10)$$

Ce résultat est à rapprocher du théorème équivalent en absence de contrôle, qui stipule que l'estimateur empirique (EE)

$$\hat{F}(y) |_{\text{EE}} = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{Y_j \leq y} = \frac{N_{00}(y) + N_{10}(y)}{n}$$

est asymptotiquement normal

$$\sqrt{n}(\hat{F}(y) - F(y)) |_{\text{EE}} \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, \sigma^2), \quad \sigma^2 = F(y)(1-F(y))$$

ce qui montre une réduction de variance apportée par la méthode VC par le facteur  $1 - \rho_I^2$ .

### 5.3.3 Estimation de quantile

On souhaite estimer le  $\alpha$ -quantile de  $Y$  en utilisant les résultats précédents sur l'estimation de la fonction de répartition de  $Y$  avec variable de contrôle. On considère la statistique d'ordre  $(Y_{(1)}, \dots, Y_{(n)})$  et les poids  $W_{(i)}$  définis par (5.3.7) associés à  $Y_{(i)}$ . Vue l'estimation (5.3.8) de la fonction de répartition, l'estimateur du  $\alpha$ -quantile est

$$\hat{Y}_{\alpha,n} = Y_{(K)}, \quad K = \inf \left\{ j, \sum_{i=1}^j W_{(i)} > \alpha \right\}$$

En utilisant des résultats sur la réduction de variance pour les méthodes de Monte Carlo par variable de contrôle, on trouve que cet estimateur est asymptotiquement normal avec la variance

$$\sqrt{n}(\hat{Y}_{\alpha,n} - y_\alpha) \mid_{VC} \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, \sigma^2), \quad \sigma^2 = \frac{\alpha(1-\alpha)}{p^2(y_\alpha)}(1 - \rho_I^2) \quad (5.3.11)$$

On rappelle que  $p$  est la densité de  $Y$  et  $\rho_I$  est le coefficient de corrélation entre  $\mathbf{1}_{Y \leq y_\alpha}$  et  $\mathbf{1}_{Z \leq z_\alpha}$ .

$$\begin{aligned} \rho_I &= \frac{\mathbb{P}(Y \leq y_\alpha, Z \leq z_\alpha) - \mathbb{P}(Y \leq y_\alpha)\mathbb{P}(Z \leq z_\alpha)}{\sqrt{(1 - \mathbb{P}(Y \leq y_\alpha))\mathbb{P}(Y \leq y_\alpha)(1 - \mathbb{P}(Z \leq z_\alpha))\mathbb{P}(Z \leq z_\alpha)}} \\ &= \frac{\mathbb{P}(Y \leq y_\alpha, Z \leq z_\alpha) - \alpha^2}{\alpha - \alpha^2} \end{aligned} \quad (5.3.12)$$

Ce résultat est à rapprocher du théorème équivalent en absence de contrôle

$$\sqrt{n}(\hat{Y}_{\alpha,n} - y_\alpha) \mid_{EE} \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, \sigma^2), \quad \sigma^2 = \frac{\alpha(1-\alpha)}{p^2(y_\alpha)}$$

ce qui montre une réduction de variance par le facteur  $1 - \rho_I^2$ . Comme attendu, plus les v.a.  $Y$  et  $Z$  sont corrélées, et meilleure est la réduction de variance. Il n'est pas facile de construire un estimateur de la variance asymptotique, car il faut pour cela estimer la densité  $p(y_\alpha)$ . Par contre, il est facile de construire un estimateur du coefficient de corrélation  $\rho_I$ , ce qui permet d'avoir une idée de la réduction de la variance. Cet estimateur est simplement le coefficient de corrélation empirique  $\hat{\rho}_I$  :

$$\hat{\rho}_I = \frac{\sum_{j=1}^n (\mathbf{1}_{Y_j \leq y} - \hat{F}_n(y))(\mathbf{1}_{Z_j \leq z_\alpha} - \hat{G}_n(z_\alpha))}{\sqrt{\sum_{j=1}^n (\mathbf{1}_{Z_j \leq z_\alpha} - \hat{G}_n(z_\alpha))^2} \sqrt{\sum_{j=1}^n (\mathbf{1}_{Y_j \leq y} - \hat{F}_n(y))^2}} \Big|_{y=\hat{Y}_{\alpha,n}}$$

avec

$$\hat{F}_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{Y_i \leq y}, \quad \hat{G}_n(z_\alpha) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{Z_i \leq z_\alpha}$$

### 5.3.4 L'estimateur avec contrôle optimal

Dans les sections précédentes nous avons arbitrairement choisi la fonction  $g(z) = \mathbf{1}_{z \leq z_\alpha}$  pour définir la variable de contrôle. Ce choix est pertinent dans le sens où il permet une

implémentation aisée et une réduction de variance substantielle, mais il n'est pas optimal. En général, la réduction de variance obtenue avec un estimateur avec contrôle de la forme (5.3.6) dépend du coefficient de corrélation entre  $\mathbf{1}_{Y \leq y}$  et le contrôle  $g(Z)$ . Le contrôle optimal, qui maximise le coefficient de corrélation, est obtenu avec la fonction [Rao 73]

$$g^*(z) = \mathbb{P}(Y \leq y | Z = z) \quad (5.3.13)$$

Cette fonction est (supposée) inconnue, car sinon, on saurait calculer analytiquement la fonction de répartition  $F(y)$  en calculant directement l'espérance de  $g^*(Z)$ , et on saurait alors déterminer numériquement le  $\alpha$ -quantile en résolvant l'équation  $F(y) = \alpha$ . Cependant, cette remarque donne l'idée de base pour la mise au point de méthodes par variable de contrôle raffinées, utilisant des approximations de la fonction de contrôle optimale  $g^*$ . Des approximations continues ont été proposées qui sont difficiles à implémenter en pratique [HJ 90]. Des approximations discrètes ont aussi été présentées, qui elles se sont révélées faciles à implémenter et très efficaces [HeN 98]. Nous décrivons maintenant cette méthode. On commence par choisir  $m + 1$  niveaux  $0 = \alpha_0 < \alpha_1 < \dots < \alpha_m = 1$ , et on note  $-\infty = z_{\alpha_0} < z_{\alpha_1} < \dots < z_{\alpha_m} = \infty$  les quantiles correspondants de  $Z$ . Les intervalles  $(z_{\alpha_{j-1}}, z_{\alpha_j}]$  vont servir de strates dans la construction d'une approximation constante par morceaux de la fonction de contrôle optimale. Cette construction est basée sur le développement suivant de la fonction de répartition de  $Y$ , qui n'est rien d'autre que la formule des probabilités totales :

$$F(y) = \sum_{j=1}^m \mathbb{P}(Y \leq y | Z \in (z_{\alpha_{j-1}}, z_{\alpha_j}]) (\alpha_j - \alpha_{j-1}) \quad (5.3.14)$$

Les quantiles de  $Z$  étant connus, on voit que l'estimation de  $F(y)$  se réduit à l'estimation des probabilités conditionnelles :

$$p_j(y) = \mathbb{P}(Y \leq y | Z \in (z_{\alpha_{j-1}}, z_{\alpha_j}]) \quad (5.3.15)$$

L'estimateur dit "Poststratified Sampling" (PS) de  $F(y)$  est

$$\hat{F}(y) = \sum_{j=1}^m \hat{p}_j(y) (\alpha_j - \alpha_{j-1})$$

où

$$\hat{p}_j(y) = \frac{\sum_{i=1}^n \mathbf{1}_{Z_i \in (z_{\alpha_{j-1}}, z_{\alpha_j}]} \mathbf{1}_{Y_i \leq y}}{\sum_{i=1}^n \mathbf{1}_{Z_i \in (z_{\alpha_{j-1}}, z_{\alpha_j}]}}$$

L'estimateur PS peut aussi bien s'écrire comme une somme pondérée des variables indicatrices  $\mathbf{1}_{Y_j \leq y}$ . La variance de l'estimateur PS est

$$\text{Var}(\hat{F}(y)) |_{\text{PS}} = \frac{1}{n} \sum_{j=1}^m (\alpha_j - \alpha_{j-1}) [p_j(y) - p_j(y)^2] + O\left(\frac{1}{n^2}\right) \quad (5.3.16)$$

En utilisant des exemples explicites gaussiens, il a été montré dans [HeN 98] que la réduction de variance optimale (celle que l'on obtient avec la fonction de contrôle  $g^*$ ) peut être pratiquement atteinte avec une approximation discrète à deux ou trois strates. En se basant sur des simulations numériques, les auteurs donnent des recommandations sur le choix du niveau  $\alpha_1 = \alpha$  pour la stratégie PS à deux strates. Ils appliquent aussi la stratégie à trois strates sur quelques exemples particuliers. Dans les prochaines sections, nous allons montrer que l'on peut en fait aller au delà de la réduction obtenue avec la fonction de contrôle optimale  $g^*(Z)$  ou avec ses approximations si on prend soin d'utiliser le modèle réduit de manière différente.

## 5.4 Estimation de quantile par stratification contrôlée

On va pousser plus loin les idées de la section précédente, pour utiliser encore plus l'existence d'un modèle réduit  $Z = f_r(X)$ . Utiliser ce modèle réduit pour estimer des quantiles n'est pas en général efficace, car un tel modèle est généralement une surface de réponse calculée pour mimer la réponse du modèle complet  $f(X)$  pour des réalisations typiques de  $X$ , et pas pour prédire la réponse de  $f(X)$  pour des réalisations exceptionnelles de  $X$ . Or c'est précisément ce que nous voulons faire lorsque nous cherchons à estimer des quantiles.

Dans cette section, nous allons utiliser le modèle réduit non pas pour approcher la réponse du modèle complet  $Y = f(X)$  dans des configurations exceptionnelles, mais pour générer un échantillon de  $X$  dans des zones où on s'attend à ce qu'il se passe quelque chose. L'idée grossière est simplement de tirer un  $X$  selon sa loi originale et de calculer  $f_r(X)$  par le modèle réduit. Si le modèle réduit prédit qu'il ne se passe rien de spécial, alors on rejette le  $X$  en question (ou plus exactement, on a tendance à le rejeter). Si le modèle réduit prédit que quelque chose de spécial arrive, alors on calcule  $f(X)$ .

On va voir dans cette section comment on peut mettre en pratique cette idée. La stratification contrôlée consiste à stratifier l'espace des valeurs prises par  $Z = f_r(X)$  en  $m$  intervalles  $I_1, \dots, I_m$ , et à forcer le nombre de réalisations de  $X$  qui sont telles que  $Z = f_r(X)$  tombent dans un intervalle  $I_j$ .

### 5.4.1 Estimation de la fonction de répartition

#### 5.4.1.1 Estimateur avec stratification contrôlée

Donnons-nous  $m + 1$  niveaux  $0 = \alpha_0 < \alpha_1 < \dots < \alpha_m = 1$ , et les quantiles de  $Z$  correspondant  $-\infty = z_{\alpha_0} < z_{\alpha_1} < \dots < z_{\alpha_m} = \infty$ . On va utiliser les intervalles  $]z_{\alpha_{j-1}}, z_{\alpha_j}]$  comme strates.

Comme on l'a vu ci-dessus, la formule des probabilités totales donne :

$$F(y) = \mathbb{P}(Y \leq y) = \sum_{j=1}^m \mathbb{P}(Y \leq y \mid Z \in ]z_{\alpha_{j-1}}, z_{\alpha_j}]) (\alpha_j - \alpha_{j-1})$$

Pour estimer  $F(y)$ , il faut donc estimer les probabilités conditionnelles

$$p_j(y) = \mathbb{P}(Y \leq y \mid Z \in ]z_{\alpha_{j-1}}, z_{\alpha_j}])$$

On se donne une suite d'entiers  $N_1, \dots, N_m$  tels que  $\sum_{j=1}^m N_j = n$ . Pour chaque  $j$ , on tire (par rejet)  $N_j$  réalisations des v.a. d'entrées  $(X_i^{(j)})_{i=1, \dots, N_j}$  telles que les sorties  $Z_i^{(j)}$  correspondantes soient dans  $]z_{\alpha_{j-1}}, z_{\alpha_j}]$ . Pour chacune de ces  $N_j$  réalisations, on calcule  $Y_i^{(j)}$ . On peut alors estimer  $p_j(y)$  par

$$\hat{p}_j(y) = \frac{1}{N_j} \sum_{i=1}^{N_j} \mathbf{1}_{Y_i^{(j)} \leq y}$$

et un estimateur de  $F(y)$  est

$$\hat{F}(y) = \sum_{j=1}^m \hat{p}_j(y) (\alpha_j - \alpha_{j-1}) \quad (5.4.1)$$

Les propriétés principales de cet estimateur sont les suivantes. L'estimateur  $\hat{F}(y)$  est sans biais

$$\mathbb{E}[\hat{F}(y)] \mid_{\text{SC}} = F(y)$$

et sa variance est

$$\text{Var}(\hat{F}(y)) \mid_{\text{SC}} = \sum_{j=1}^m \frac{(\alpha_j - \alpha_{j-1})^2}{N_j} [p_j(y) - p_j(y)^2] \quad (5.4.2)$$

Si  $m$  est fixé, si les  $\beta_j$  sont des réels positifs tels que  $\sum_{j=1}^m \beta_j = 1$ , et si on note  $N_j = [n\beta_j]$ , alors l'estimateur  $\hat{F}(y)$  est asymptotiquement normal

$$\sqrt{n}(\hat{F}(y) - F(y)) \mid_{\text{SC}} \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, \sigma^2), \quad \sigma^2 = \sum_{j=1}^m \frac{(\alpha_j - \alpha_{j-1})^2}{\beta_j} [p_j(y) - p_j(y)^2] \quad (5.4.3)$$

#### 5.4.1.2 Estimation de la fonction de répartition

On cherche ici à estimer la fonction de répartition  $F(y)$  pour tout  $y \in \mathbb{R}$ .

Dans le cas où la variable de contrôle  $Z$  est indépendante de  $Y$  (c'est-à-dire qu'elle ne contrôle rien du tout), on obtient  $p_j(y) = F(y)$  et

$$\text{Var}(\hat{F}(y)) \mid_{\text{SC}} = \left[ \sum_{j=1}^m \frac{(\alpha_j - \alpha_{j-1})^2}{N_j} \right] \times [F(y) - F(y)^2] \quad (5.4.4)$$

Si  $N_j = [(\alpha_j - \alpha_{j-1})n]$ , alors on trouve que, modulo les erreurs d'arrondi, la variance de l'estimateur

$$\text{Var}(\hat{F}(y)) \mid_{\text{SC}} = \frac{1}{n} [F(y) - F(y)^2]$$

est égale à la variance de l'estimateur empirique. C'est en particulier le cas si on choisit une répartition régulière  $\alpha_j = j/m$  et  $N_j = n/m$ .

Dans le cas où la variable de contrôle  $Z$  est une fonction croissante de  $Y$  (c'est-à-dire qu'elle la contrôle totalement), alors on obtient, si  $y \in ]y_{\alpha_{j-1}}, y_{\alpha_j}]$ ,

$$\text{Var}(\hat{F}(y)) |_{\text{SC}} = \frac{(\alpha_j - \alpha_{j-1})^2}{N_j} [p_j(y) - p_j(y)^2] \leq \frac{(\alpha_j - \alpha_{j-1})^2}{4N_j}$$

Si  $\alpha_j = j/m$  et  $N_j = n/m$ , alors

$$\text{Var}(\hat{F}(y)) |_{\text{SC}} \leq \frac{1}{4mn}$$

On a donc réduit la variance de l'estimateur d'un facteur  $1/m$ . La méthode est donc efficace pour reconstruire la fonction de répartition de  $Y$ .

#### 5.4.1.3 Estimation de la queue de la fonction de répartition

On cherche ici à estimer la queue de la fonction de répartition  $F(y)$ , disons dans la zone où  $F(y) \simeq 1 - \delta$  avec  $\delta$  proche de 0.

Si  $Z$  et  $Y$  sont corrélées positivement, on a alors intérêt à mettre plus de points dans la queue de distribution de la v.a. de contrôle  $Z$ , afin de renforcer le nombre de réalisations potentiellement intéressantes.

Prenons par exemple  $m = 4$ ,  $\alpha_1 = 1/2$ ,  $\alpha_2 = 1 - 2\delta$ ,  $\alpha_3 = 1 - \delta$ ,  $N_j = n/4$ , c'est-à-dire qu'on tire :

- $n/4$  réalisations telles que  $Z_i \leq z_{\alpha_1}$ ,
- $n/4$  réalisations telles que  $z_{\alpha_1} < Z_i \leq z_{\alpha_2}$ ,
- $n/4$  réalisations telles que  $z_{\alpha_2} < Z_i \leq z_{\alpha_3}$ ,
- $n/4$  réalisations telles que  $z_{\alpha_3} < Z_i$ .

On a donc  $n/2$  points dans la queue de distribution de  $Z$ , pour lesquels  $z_{2\alpha-1} < Z$ .

Si  $Z$  et  $Y$  sont indépendantes, alors Eq. (5.4.4) montre que cette stratégie nous fait perdre un facteur 2 dans la variance de l'estimateur :

$$\text{Var}(\hat{F}(y)) |_{\text{SC}} \simeq \frac{2\delta}{n}$$

par rapport au cas sans contrôle :

$$\text{Var}(\hat{F}(y)) |_{\text{EE}} \simeq \frac{\delta}{n}$$

Si  $Z$  et  $Y$  sont fortement corrélées, suffisamment pour que les réalisations telles que  $Z \leq z_{1-2\delta}$  et  $Y > y_{1-\delta}$  soient négligeables, alors Eq. (5.4.2) montre que cette stratégie nous fait

gagner un facteur  $\delta$  dans la variance de l'estimateur :

$$\text{Var}(\hat{F}(y)) |_{\text{SC}} \simeq \frac{\delta^2}{n}$$

Le gain est donc, dans le cas où les variables  $Y$  et  $Z$  sont corrélées, très substantiel.

## 5.4.2 Estimation d'un quantile

### 5.4.2.1 Mise en oeuvre

On est intéressé par la détection d'un  $\alpha$ -quantile de  $Y$ , avec  $\alpha$  proche de 1. Un estimateur du  $\alpha$ -quantile de  $Y$  est :

$$\hat{Y}_{\alpha,n} |_{\text{SC}} = \inf \left\{ y, \hat{F}(y) > \alpha \right\}$$

où  $\hat{F}(y)$  est l'estimateur de la fonction de répartition que l'on vient de décrire. La définition est correcte car  $\hat{F}(y)$  est une fonction croissante de  $y$ , qui est nulle pour  $y$  assez petit et égale à 1 pour  $y$  assez grand.

L'estimateur  $\hat{Y}_{\alpha,n}$  est asymptotiquement normal

$$\sqrt{n}(\hat{Y}_{\alpha,n} - y_\alpha) |_{\text{SC}} \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, \sigma^2), \quad \sigma^2 = \frac{\sum_{j=1}^m \frac{(\alpha_j - \alpha_{j-1})^2}{\beta_j} [p_j(y_\alpha) - p_j(y_\alpha)^2]}{p^2(y_\alpha)}$$

où  $p$  est la densité de  $Y$ .

Si  $Z$  et  $Y$  sont corrélées positivement, on a alors intérêt à mettre plus de points dans la queue de distribution de la v.a. de contrôle  $Z$ , afin de renforcer le nombre de réalisations potentiellement intéressantes. La stratification contrôlée en quatre strates décrite page 167 semble un bon compromis. Du moins, elle apparaît performante sur les quelques exemples que nous avons étudiés.

Nous donnons dans le prochain paragraphe les résultats pour quatre exemples qui nous ont semblé typiques de situations réalistes.

### 5.4.2.2 Simulations

#### Exemple 1 : fonction 1D

On considère la situation suivante.  $X$  est supposée être une v.a. gaussienne centrée réduite, et les fonctions  $f$  et  $f_r$  sont données par

$$f_r(x) = x^2, \quad f(x) = 0.95x^2[1 + 0.5 \cos(10x) + 0.5 \cos(20x)] \quad (5.4.5)$$

et sont dessinées sur la figure 5.4.1a. Les quantiles de la v.a. de contrôle  $Z = f_r(X)$  sont calculables analytiquement :

$$z_\alpha = \left[ \Phi^{-1} \left( \frac{1 + \alpha}{2} \right) \right]^2$$

Ceux de  $Y$  ne le sont pas, comme on peut le voir en traçant la densité de probabilité de  $Y$  (figure 5.4.1b) obtenue par une simulation de Monte Carlo intensive (avec  $5 \times 10^7$  réalisations).

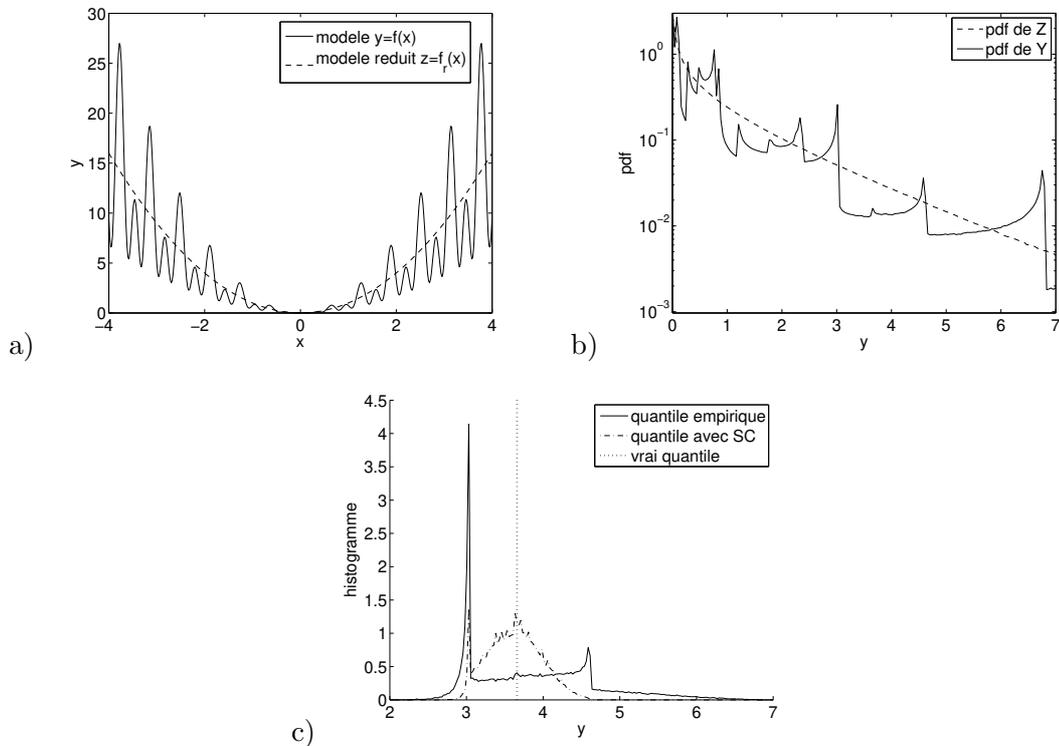


FIG. 5.4.1 – Figure a : graphes des fonctions  $f$  et  $f_r$  données par (5.4.5). Figure b : pdf de  $Y = f(X)$  et  $Z = f_r(X)$  pour  $X \sim \mathcal{N}(0,1)$ , obtenues avec  $5 \times 10^7$  tirages de Monte Carlo. Figure c : Estimations du  $\alpha$ -quantile pour la v.a.  $Y = f(X)$  avec  $\alpha = 95\%$  à partir d'un échantillon de taille  $n = 200$ . Les histogrammes des estimateurs sont tracés à partir de  $10^4$  expériences.

On souhaite déterminer le  $\alpha$ -quantile de  $Y$ . On dispose d'un nombre limité  $n$  d'appels à la fonction  $f$ , mais d'un nombre quasi-illimité d'appels à la fonction réduite  $f_r$ . On utilise la méthode de stratification contrôlée avec les paramètres décrits page 167, et on compare le quantile obtenu avec le quantile empirique (figure 5.4.1c). On peut voir d'une part que le quantile empirique a beaucoup de mal à prédire la valeur du vrai quantile, tandis que l'estimateur avec stratification contrôlée apparaît plus performant. En utilisant des simulations intensives, nous avons déterminé le 0.95-quantile de  $Y$  :  $y_{0.95} = 3.66$ , ainsi que le coefficient de corrélation entre  $Y$  et  $Z$  :  $\rho = 0.84$ . La réduction de variance de la méthode par stratification contrôlée est lié au coefficient  $\rho_I$ , qui peut aussi être évalué par les simulations et l'équation (5.3.12) :  $\rho_I = 0.52$ .

### Exemple 2 : fonction 2D

On effectue le même travail dans le cas où  $X = (X_1, X_2)$  est un couple de v.a. normales centrées réduites indépendantes et les fonctions  $f$  et  $f_r$  sont données par :

$$f_r(x_1, x_2) = x_1^2 x_2, \quad (5.4.6)$$

$$f(x_1, x_2) = 0.95x_1^2[1 + 0.5 \cos(10x_1) + 0.5 \cos(20x_1)] \\ \times x_2[1 + 0.4 \cos(7x_2) + 0.3 \cos(14x_2)] \quad (5.4.7)$$

Ces fonctions et leur densité de probabilité sont illustrées en figure 5.4.2 (a, b, c).

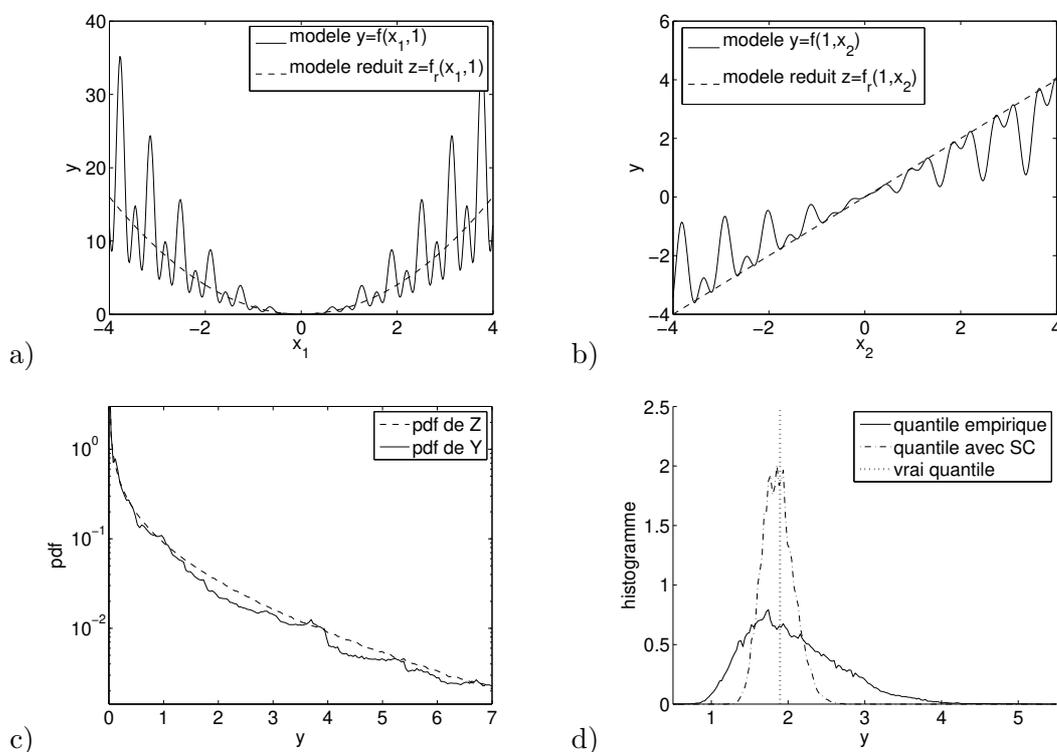


FIG. 5.4.2 – Figures a,b : graphes en coupe des fonctions  $f(x_1, x_2)$  et  $f_r(x_1, x_2)$  données par (5.4.6-5.4.7). Figure c : pdf de  $Y = f(X_1, X_2)$  et  $Z = f_r(X_1, X_2)$  pour  $X_1, X_2 \sim \mathcal{N}(0, 1)$ , obtenues avec  $5 \times 10^7$  tirages de Monte Carlo. Figure d : Estimations du  $\alpha$ -quantile pour la v.a.  $Y = f(X_1, X_2)$  avec  $\alpha = 95\%$  à partir d'un échantillon de taille  $n = 200$ . Les histogrammes des estimateurs sont tracés à partir de  $10^4$  expériences.

Ici encore, la méthode par stratification contrôlée réduit de manière significative la variance de l'estimateur du quantile, par rapport à l'estimateur empirique, comme on peut le voir sur la figure 5.4.2d.

### Exemple 3 : fonction d'Ishigami

On effectue le même travail dans le cas où  $X = (X_1, X_2, X_3)$  est une famille de trois v.a. uniformes sur  $[-\pi, \pi]$  et les fonctions  $f$  et  $f_r$  sont données par :

$$f_r(x_1, x_2, x_3) = 1.908 + 1.727x_1 + 2.059x_2^2 - 0.2756x_1^3 - 0.2663x_2^4 + 0.2499x_3^2x_1, \quad (5.4.8)$$

$$f(x_1, x_2, x_3) = \sin(x_1) + 7 \sin(x_2)^2 + 0.1x_3^4 \sin(x_1) \quad (5.4.9)$$

Ici, la fonction  $f$  est la fonction d'Ishigami, bien connue en analyse de sensibilité [HS 96], et illustrée en figure 5.4.3 (a, b, c, d).  $f_r$  est une surface de réponse polynomiale à cinq monômes (l'intercept ne jouant aucun rôle dans le contrôle). Elle a été ajustée par régression linéaire multiple à partir d'une base de 10000 calculs de  $f$ . Seuls les monômes dont les coefficients ont été jugés significativement non nuls par le test de Student (t-values) ont été conservés. On constate que les termes de  $f_r$  correspondent à ceux du développement limité à l'ordre 4 de la fonction  $f$ , et ainsi que la surface de réponse ne reproduit pas toute la variabilité de  $f$  (figure 5.4.3 a, b, c, d).

Les capacités d'approximation de la surface de réponse peuvent être mesurées à l'aide du coefficient de détermination de la régression :  $R^2 = 0.75$ . Par validation croisée (la base des 10000 points est divisée en 100 intervalles), on vérifie que le  $R^2$  en prédiction est du même ordre de grandeur :  $R_{pred}^2 = 0.74$ . Dans la méthode de stratification contrôlée, c'est la valeur du coefficient de corrélation entre  $f(X)$  et  $f_r(X)$  qui est importante ; ici on a  $\rho = 0.86$ , ce qui montre une corrélation forte.

Le  $\alpha$ -quantile de  $Z = f_r(X)$  est  $z_\alpha \simeq 8.51$ , assez loin du  $\alpha$ -quantile de  $Y = f(X)$  qui est  $y_\alpha \simeq 9.30$  (voir figure 5.4.3e). La surface de réponse ne peut pas être utilisée ici pour estimer directement le quantile, mais elle peut être efficacement utilisée avec la stratification contrôlée : celle-ci réduit de manière significative la variance de l'estimateur du quantile, par rapport à l'estimateur empirique, comme on peut le voir sur la figure 5.4.3e.

#### Exemple 4 : fonction de Morris

On effectue le même travail dans le cas où  $X = (X_i)_{i=1, \dots, 10}$  est une famille de dix v.a. uniformes sur  $[0, 1]$  et la fonction  $f$  est une simplification de la fonction de Morris, bien connue

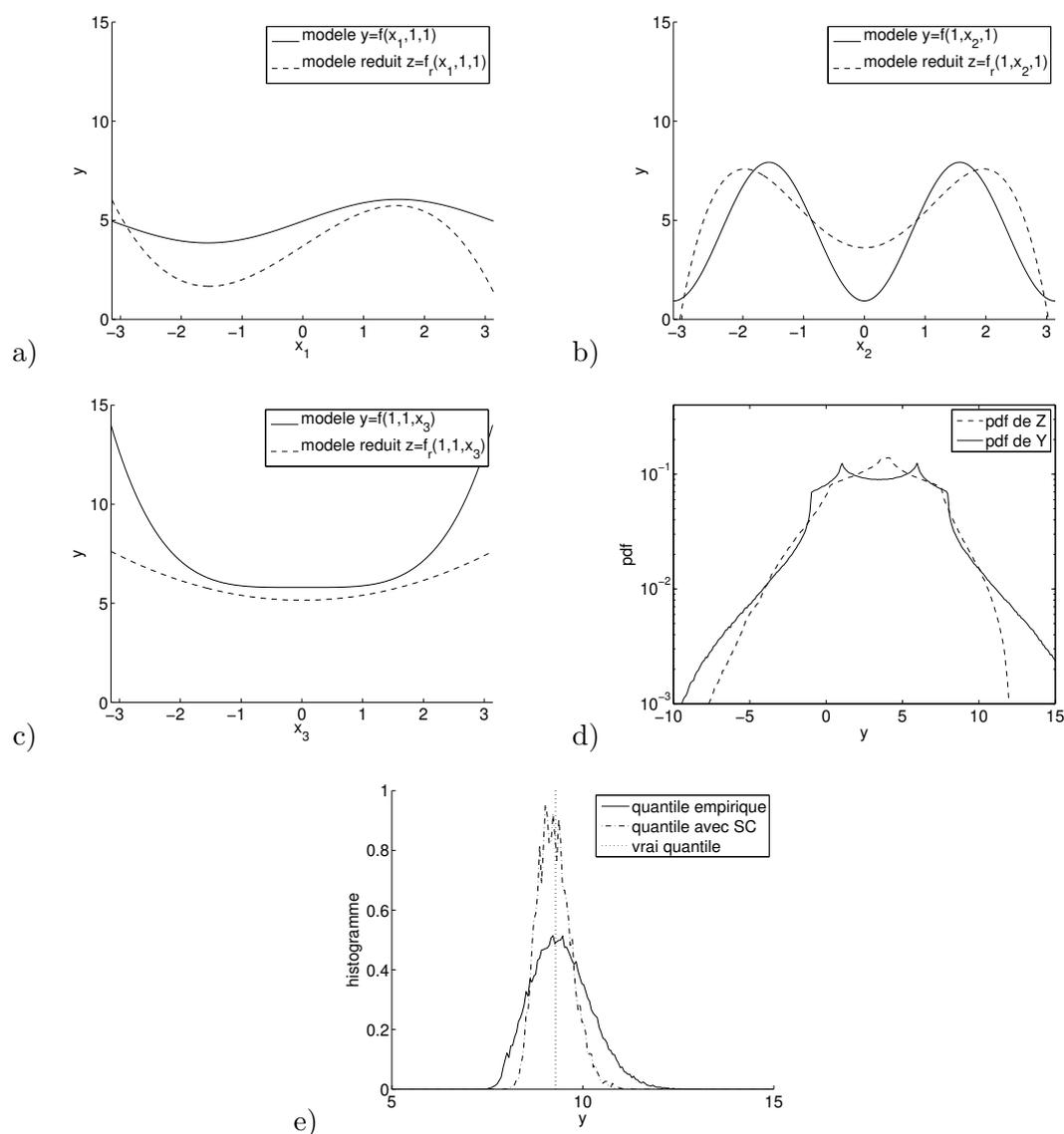


FIG. 5.4.3 – Figures a,b,c : graphes en coupe des fonctions  $f(x_1, x_2, x_3)$  et  $f_r(x_1, x_2, x_3)$  données par (5.4.8-5.4.9). Figure d : pdf de  $Y = f(X_1, X_2, X_3)$  et  $Z = f_r(X_1, X_2, X_3)$  pour  $X_1, X_2, X_3 \sim \mathcal{U}(-\pi, \pi)$ , obtenues avec  $5 \times 10^7$  tirages de Monte Carlo. Figure e : Estimations du  $\alpha$ -quantile pour la v.a.  $Y = f(X_1, X_2, X_3)$  avec  $\alpha = 95\%$  à partir d'un échantillon de taille  $n = 200$ . Les histogrammes des estimateurs sont tracés à partir de  $10^4$  expériences.

en analyse de sensibilité [Mor 91], donnée par :

$$f(x) = \sum_{i=1}^{10} w_i(x) + \sum_{1 \leq i < j \leq 6} w_i(x)w_j(x) + \sum_{1 \leq i < j < k \leq 5} w_i(x)w_j(x)w_k(x) + \sum_{1 \leq i < j < k < l \leq 4} w_i(x)w_j(x)w_k(x)w_l(x) \quad (5.4.10)$$

$$w_i(x) = \begin{cases} 2 \left( \frac{1.1x_i}{x_i + 0.1} - 0.5 \right) & \text{si } i = 3, 5, 7 \\ 2(x_i - 0.5) & \text{sinon} \end{cases} \quad (5.4.11)$$

La fonction  $f_r$  est une surface de réponse polynomiale de degré deux, ajustée par régression linéaire multiple à partir d'une base de 10000 calculs de  $f$ . Deux surfaces de réponse sont testées, une pas très bien ajustée (avec  $R^2 = 0.7$  et  $\rho = 0.82$ ), et l'autre de bonne qualité avec ( $R^2 = 0.9$  et  $\rho = 0.95$ ). Leur densité de probabilité est comparée à celle de  $f$  en figure 5.4.4 (a, b).

On compare sur la figure 5.4.4 (c, d) les résultats de la stratification contrôlée avec ces deux surfaces de réponse. Une fois de plus, la stratification contrôlée réduit de manière significative la variance de l'estimateur du quantile. Cette réduction est d'autant plus importante que le coefficient de corrélation  $\rho$  entre  $f(X)$  et  $f_r(X)$  est élevé.

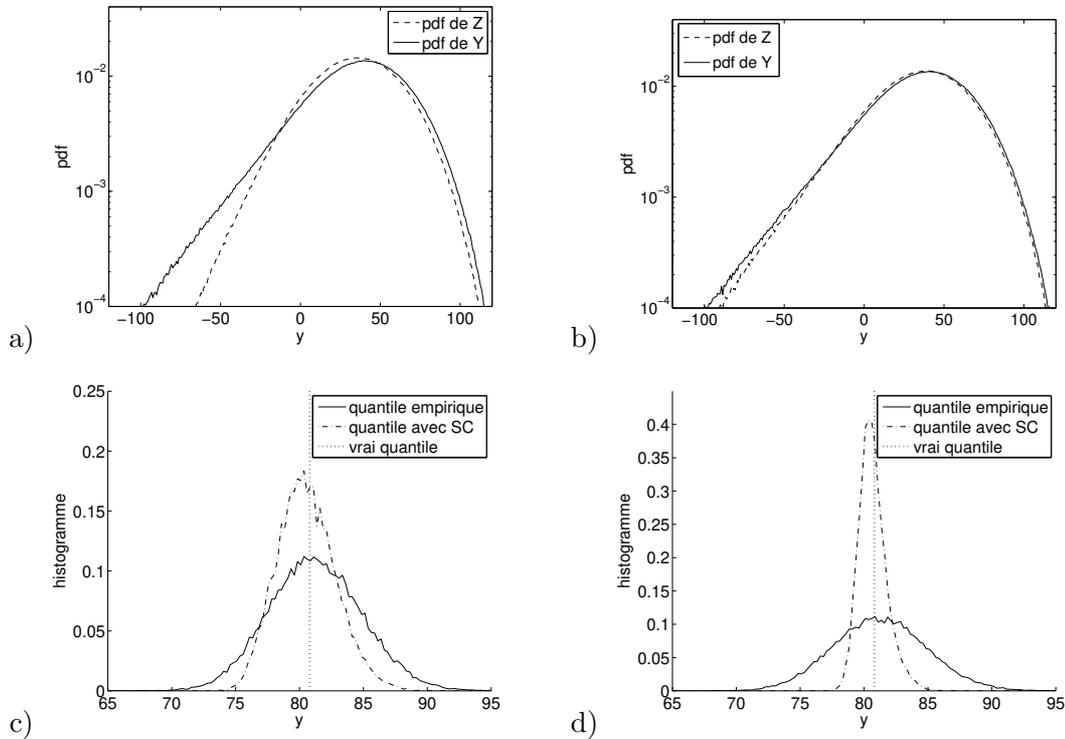


FIG. 5.4.4 – En haut : pdf de  $Y = f(X)$  et  $Z = f_r(X)$  pour  $X = (X_i)_{i=1,\dots,10}$  et  $X_i \sim \mathcal{U}(0, 1)$ , avec une surface de réponse de  $R^2 = 0.7$  (a) et  $R^2 = 0.9$  (b). En bas : Estimations du  $\alpha$ -quantile pour la v.a.  $Y = f(X)$  avec  $\alpha = 95\%$  à partir d'un échantillon de taille  $n = 200$ . Les histogrammes des estimateurs sont tracés à partir de  $10^4$  expériences. On utilise une surface de réponse avec  $R^2 = 0.7$  (resp.  $R^2 = 0.9$ ) comme variable de contrôle sur la figure c (resp. d).

### 5.4.3 Stratification contrôlée adaptative

#### 5.4.3.1 Choix optimal du nombre de simulations par strates

Considérons l'estimateur par stratification contrôlée décrit dans la section 5.4.1.1 et reprenons les mêmes notations. Fixons  $y \in \mathbb{R}$ . L'estimateur (5.4.1) de  $F(y)$  est asymptotiquement

normal et sa variance réduite  $\sigma^2$  est donnée par (5.4.3) :

$$\sigma^2 = \sum_{j=1}^m \frac{(\alpha_j - \alpha_{j-1})^2}{\beta_j} [p_j(y) - p_j(y)^2]$$

En fait, si on néglige les petites erreurs dues aux parties entières,  $\sigma^2/n$  est la variance de l'estimateur  $\hat{F}(y)$  pour tout  $n$  en vertu de (5.4.2).

Si on se fixe le nombre  $m$  et les niveaux  $(\alpha_j)_{j=0,\dots,m}$  des strates, ainsi que le nombre total de simulations  $n$ , alors il est possible de choisir le nombre de simulations  $[\beta_j n]$  par strate pour minimiser la variance de l'estimateur. En posant  $q_j = (\alpha_j - \alpha_{j-1})^2 [p_j(y) - p_j(y)^2]$ , le problème de minimisation

$$\operatorname{argmin} \left\{ \sum_{j=1}^m \frac{q_j}{\beta_j} \right\} \quad \text{sous la contrainte} \quad \beta_j \geq 0, \quad \sum_{j=1}^m \beta_j = 1$$

a pour solution :

$$\beta_j = \frac{q_j^{1/2}}{\sum_{l=1}^m q_l^{1/2}}$$

Avec ce choix optimal, la variance réduite de l'estimateur  $\hat{F}(y)$  est

$$\sigma^2 = \left\{ \sum_{j=1}^m (\alpha_j - \alpha_{j-1}) [p_j(y) - p_j(y)^2]^{1/2} \right\}^2 \quad (5.4.12)$$

Noter que le choix optimal des  $\beta_j$  dépend en général de  $y$ , et qu'on ne peut donc pas proposer de choix qui soit optimal pour l'estimation complète de la fonction de répartition. On peut cependant voir que :

- 1) Si le contrôle est faible, alors  $p_j(y)$  dépend faiblement de  $y$ , et le choix optimal est alors  $\beta_j = \alpha_j - \alpha_{j-1}$ .
- 2) Si le contrôle est fort, alors il faut mettre plus de simulations dans les strates  $]z_{\alpha_{j-1}}, z_{\alpha_j}]$  autour de  $F(y)$ . En effet, supposons par exemple un contrôle très fort, dans le sens où  $Z = \phi(Y)$  est une fonction croissante de  $Y$ . Alors

$$p_j(y) = \mathbb{P}(Y \leq y | Z \in ]z_{\alpha_{j-1}}, z_{\alpha_j}]) = \mathbb{P}(Y \leq y \leq |Y \in ]\phi^{-1}(z_{\alpha_{j-1}}), \phi^{-1}(z_{\alpha_j})])$$

est égal à 1 si  $z_{\alpha_j} \leq \phi(y)$  (i.e.  $\alpha_j \leq F(y)$ ) et à 0 si  $z_{\alpha_{j-1}} > \phi(y)$  (i.e.  $\alpha_{j-1} > F(y)$ ). Dans les deux cas,  $q_j$ , et donc  $\beta_j$ , est nul, et il faut mettre toutes les simulations dans la strate  $]z_{\alpha_{j_0-1}}, z_{\alpha_{j_0}}]$  pour laquelle  $\alpha_{j_0-1} < F(y) \leq \alpha_{j_0}$ . Bien sûr, le contrôle très fort supposé ici est peu réaliste, mais cet exemple illustre bien comment varie la répartition optimale des simulations dans les strates.

### 5.4.3.2 Estimateur adaptatif de la fonction de répartition

Nous savons maintenant qu'il existe un choix optimal pour la répartition des  $n$  simulations sur les  $m$  strates. Cette répartition dépend des  $p_j(y)$ , qui sont les quantités à estimer. On peut alors proposer une procédure adaptative :

1) On commence par appliquer la procédure de stratification contrôlée avec  $\tilde{n} = n^\gamma$  simulations,  $\gamma \in ]0, 1[$ , et avec un choix a priori des  $\beta_j$ , et on obtient ainsi une première estimation des  $p_j(y)$  :

$$\tilde{p}_j(y) = \frac{1}{[\beta_j \tilde{n}]} \sum_{i=1}^{[\beta_j \tilde{n}]} \mathbf{1}_{Y_i^{(j)} \leq y}$$

2) On estime la proportion idéale de simulations par strates :

$$\tilde{\beta}_j = \frac{(\alpha_j - \alpha_{j-1}) [\tilde{p}_j(y) - \tilde{p}_j(y)^2]^{1/2}}{\sum_{l=1}^m (\alpha_l - \alpha_{l-1}) [\tilde{p}_l(y) - \tilde{p}_l(y)^2]^{1/2}}$$

3) On effectue les  $n - \tilde{n}$  dernières simulations en affectant le nombre de simulations nécessaires dans chaque strate pour atteindre le nombre optimal estimé  $[n\tilde{\beta}_j]$ .

4) On peut alors estimer  $p_j(y)$  et  $F(y)$  par

$$\hat{p}_j(y) = \frac{1}{[\tilde{\beta}_j n]} \sum_{i=1}^{[\tilde{\beta}_j n]} \mathbf{1}_{Y_i^{(j)} \leq y}, \quad \hat{F}(y) = \sum_{j=1}^m \hat{p}_j(y) (\alpha_j - \alpha_{j-1})$$

L'estimateur  $\hat{F}(y)$  obtenu par stratification contrôlée adaptative (SCA) est asymptotiquement normal

$$\sqrt{n} \left( \hat{F}(y) - F(y) \right) \Big|_{\text{SCA}} \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, \sigma^2), \quad \sigma^2 = \left\{ \sum_{j=1}^m (\alpha_j - \alpha_{j-1}) [p_j(y) - p_j(y)^2]^{1/2} \right\}^2 \quad (5.4.13)$$

L'expression de la variance réduite  $\sigma^2$  est donc la même que celle de la variance réduite (5.4.12) de l'estimateur lorsqu'on choisit a priori la répartition optimale des  $\beta_j$ . La différence est que, dans la cas de l'estimateur adaptatif, la variance  $\sigma^2/n$  est atteinte asymptotiquement par  $\hat{F}(y)$  lorsque  $n \rightarrow \infty$ , alors que pour l'estimateur optimal, la variance est  $\sigma^2/n$  pour tout  $n$ .

### 5.4.3.3 Réduction de variance

On va mener ici une analyse de la réduction de variance apportée par la méthode de stratification contrôlée adaptative ou non dans le cas de  $m = 2$  strates.

Dans le cas de la méthode PS, ou dans le cas de la stratification contrôlée en imposant  $\beta_j = \alpha_j - \alpha_{j-1}$ , la variance réduite dans (5.4.3) est

$$\sigma^2 |_{\text{PS}} = \alpha^{-1} \mathbb{P}(Y \leq y, Z \leq z_\alpha) \mathbb{P}(Y > y, Z \leq z_\alpha) + (1 - \alpha)^{-1} \mathbb{P}(Y \leq y, Z > z_\alpha) \mathbb{P}(Y > y, Z > z_\alpha)$$

Dans le cas de la stratification contrôlée adaptative, l'expression de la variance réduite  $\sigma^2$  dans (5.4.13) est

$$\begin{aligned} \sigma^2 |_{\text{SCA}} = & \left[ \mathbb{P}(Y \leq y, Z \leq z_\alpha)^{1/2} \mathbb{P}(Y > y, Z \leq z_\alpha)^{1/2} \right. \\ & \left. + \mathbb{P}(Y \leq y, Z > z_\alpha)^{1/2} \mathbb{P}(Y > y, Z > z_\alpha)^{1/2} \right]^2 \end{aligned}$$

En fonction du coefficient de corrélation  $\rho_I$  entre  $\mathbf{1}_{Y \leq y}$  et  $\mathbf{1}_{Z \leq z_\alpha}$ , défini par (5.3.10), on peut écrire :

$$\begin{aligned} \sigma^2 |_{\text{PS}} = & F(y)(1 - F(y)) \\ & \times \left\{ \alpha \left[ 1 + \rho_I \left( \frac{(1 - \alpha)(1 - F(y))}{\alpha F(y)} \right)^{1/2} \right] \left[ 1 - \rho_I \left( \frac{(1 - \alpha)F(y)}{\alpha(1 - F(y))} \right)^{1/2} \right] \right. \\ & \left. + (1 - \alpha) \left[ 1 + \rho_I \left( \frac{\alpha F(y)}{(1 - \alpha)(1 - F(y))} \right)^{1/2} \right] \left[ 1 - \rho_I \left( \frac{\alpha(1 - F(y))}{(1 - \alpha)F(y)} \right)^{1/2} \right] \right\} \\ \sigma^2 |_{\text{SCA}} = & F(y)(1 - F(y)) \\ & \times \left\{ \alpha \left[ 1 + \rho_I \left( \frac{(1 - \alpha)(1 - F(y))}{\alpha F(y)} \right)^{1/2} \right]^{1/2} \left[ 1 - \rho_I \left( \frac{(1 - \alpha)F(y)}{\alpha(1 - F(y))} \right)^{1/2} \right]^{1/2} \right. \\ & \left. + (1 - \alpha) \left[ 1 + \rho_I \left( \frac{\alpha F(y)}{(1 - \alpha)(1 - F(y))} \right)^{1/2} \right]^{1/2} \left[ 1 - \rho_I \left( \frac{\alpha(1 - F(y))}{(1 - \alpha)F(y)} \right)^{1/2} \right]^{1/2} \right\}^2 \end{aligned}$$

Pour pouvoir comparer la réduction de variance apportée par la méthode de variable de contrôle et celles apportées par les méthodes de stratification contrôlée, il faut comparer ces expressions avec (5.3.9). Pour rendre cette comparaison plus explicite, on peut par exemple supposer que  $\rho_I$  est petit, et développer chaque expression en  $\rho_I$ . On trouve

$$\sigma^2 |_{\text{VC}} = F(y)(1 - F(y)) [1 - \rho_I^2] \quad (5.4.14)$$

$$\sigma^2 |_{\text{PS}} = F(y)(1 - F(y)) [1 - \rho_I^2] \quad (5.4.15)$$

$$\sigma^2 |_{\text{SCA}} = F(y)(1 - F(y)) \left[ 1 - \frac{\rho_I^2}{8F(y)(1 - F(y))} + O(\rho_I^3) \right] \quad (5.4.16)$$

alors que la variance réduite pour l'estimateur empirique usuel est  $F(y)(1 - F(y))$ . Cela montre que les méthodes par variable de contrôle et par stratification contrôlée adaptative apportent à peu près la même réduction de variance lorsqu'on cherche à estimer la fonction de répartition autour de la médiane  $F(y) \sim 1/2$  (en fait, la méthode SCA apporte une réduction de variance par un facteur  $\sim 1 - \rho_I^2/2$  si  $F(y) \sim 1/2$ ). Mais, lorsqu'on cherche à estimer les queues de la fonction de répartition  $F(y) \sim 0$  ou  $1$ , alors la méthode par stratification contrôlée adaptative apporte une bien meilleure réduction de variance.

Les expressions que l'on vient d'obtenir permettent aussi d'orienter le choix a priori de  $\alpha$ . En effet, la réduction de variance sera d'autant plus grande que le coefficient de corrélation  $\rho_I$

est grand. Si on suppose par exemple que  $Z = \phi(Y)$  est une fonction croissante de  $Y$ , ce qui représente le cas d'un contrôle très fort, alors on trouve que

$$\rho_I = \begin{cases} \left[ \frac{\alpha(1 - F(y))}{(1 - \alpha)F(y)} \right]^{1/2} & \text{si } z_\alpha < \phi(y) \\ \left[ \frac{(1 - \alpha)F(y)}{\alpha(1 - F(y))} \right]^{1/2} & \text{si } z_\alpha \geq \phi(y) \end{cases}$$

En tant que fonction de  $\alpha$ , cette fonction est maximale lorsque  $\alpha = F(y)$ . Cela montre que, si on veut estimer la fonction de répartition autour d'un certain  $y$ , alors on a intérêt à choisir  $\alpha = F(y)$ .

Pour finir cette section, nous allons effectuer une nouvelle analyse asymptotique de la réduction de variance pour les méthodes VC, PS, SC et SCA dans le cas où le nombre  $m$  de strates est grand. Dans la méthode PS et dans la méthode SC avec la distribution  $\beta_j = \alpha_j - \alpha_{j-1}$ , la variance réduite est (5.4.3). Si la probabilité conditionnelle  $\mathbb{P}(Y \leq y|Z)$  a une densité continue  $g^*$  par rapport à la mesure de Lebesgue, alors (5.4.3) est une somme de Riemann qui a la limite suivante quand  $m \rightarrow \infty$

$$\sigma^2 |_{\text{PS}} = \mathbb{E} [\mathbb{P}(Y \leq y|Z) - \mathbb{P}(Y \leq y|Z)^2] = F(y) - \mathbb{E} [\mathbb{P}(Y \leq y|Z)^2]$$

Il s'agit en fait de la variance réduite de l'estimateur CV avec la fonction de contrôle optimale  $g^*$  définie par (5.3.13). Dans la méthode SCA, l'expression de la variance réduite  $\sigma^2$  est (5.4.13) qui a la limite suivante quand  $m \rightarrow \infty$

$$\sigma^2 |_{\text{SCA}} = \mathbb{E} \left[ (\mathbb{P}(Y \leq y|Z) - \mathbb{P}(Y \leq y|Z)^2)^{1/2} \right]^2$$

L'inégalité de Cauchy-Schwarz montre alors clairement que la réduction de variance est plus importante pour la méthode SCA que pour la méthode CV utilisant la fonction de contrôle optimale  $g^*$ .

## 5.4.4 Stratification contrôlée adaptative pour l'estimation d'un quantile

### 5.4.4.1 Mise en oeuvre

On cherche ici à déterminer le  $\alpha$ -quantile de  $Y$ , avec  $\alpha$  proche de 1. En reprenant les idées de la méthode d'estimation adaptative pour la fonction de répartition qu'on vient de décrire, on propose la procédure adaptative suivante :

1) On commence par appliquer la procédure de stratification contrôlée avec  $\tilde{n} = n^\gamma$  simulations,  $\gamma < 1$ , et avec un choix a priori des  $\beta_j$ , et on obtient ainsi une première estimation des  $p_j(y)$  :

$$\tilde{p}_j(y) = \frac{1}{[\beta_j \tilde{n}]} \sum_{i=1}^{[\beta_j \tilde{n}]} \mathbf{1}_{Y_i^{(j)} \leq y}$$

de la fonction de répartition  $\tilde{F}(y)$  :

$$\tilde{F}(y) = \sum_{j=1}^m (\alpha_j - \alpha_{j-1}) \tilde{p}_j(y)$$

et du quantile  $y_\alpha$  :

$$\tilde{Y}_\alpha = \inf \left\{ y, \tilde{F}(y) > \alpha \right\}$$

2) On estime la proposition idéale de simulations par strates pour le quantile estimé  $\tilde{Y}_\alpha$  :

$$\tilde{\beta}_j = \frac{(\alpha_j - \alpha_{j-1}) \left[ \tilde{p}_j(\tilde{Y}_\alpha) - \tilde{p}_j(\tilde{Y}_\alpha)^2 \right]^{1/2}}{\sum_{l=1}^m (\alpha_l - \alpha_{l-1}) \left[ \tilde{p}_l(\tilde{Y}_\alpha) - \tilde{p}_l(\tilde{Y}_\alpha)^2 \right]^{1/2}} \quad (5.4.17)$$

3) On effectue les  $n - \tilde{n}$  dernières simulations en affectant le nombre de simulations nécessaires dans chaque strate pour atteindre le nombre idéal estimé  $[\tilde{\beta}_j n]$ .

4) On peut alors estimer  $p_j(y)$  et  $F(y)$  par

$$\hat{p}_j(y) = \frac{1}{[\tilde{\beta}_j n]} \sum_{i=1}^{[\tilde{\beta}_j n]} \mathbf{1}_{Y_i^{(j)} \leq y}, \quad \hat{F}(y) = \sum_{j=1}^m \hat{p}_j(y) (\alpha_j - \alpha_{j-1})$$

et le quantile  $y_\alpha$  par

$$\hat{Y}_{\alpha, n} = \inf \left\{ y, \hat{F}(y) > \alpha \right\}$$

L'estimateur  $\hat{Y}_{\alpha, n}$  est asymptotiquement normal

$$\sqrt{n}(\hat{Y}_{\alpha, n} - y_\alpha) \Big|_{\text{SCA}} \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, \sigma^2), \quad \sigma^2 = \frac{\sum_{j=1}^m (\alpha_j - \alpha_{j-1}) \left[ p_j(y_\alpha) - p_j(y_\alpha)^2 \right]^{1/2}}{p^2(y_\alpha)}$$

où  $p$  est la densité de  $Y$ .

Pour conclure, on a donc obtenu que la variance réduite  $\sigma^2$  est, pour l'estimateur empirique usuel :

$$\sigma^2 \Big|_{\text{EE}} = \frac{\alpha(1 - \alpha)}{p^2(y_\alpha)}$$

pour l'estimateur PS ou pour l'estimateur VC avec la distribution  $\beta_j = \alpha_j - \alpha_{j-1}$  (voir (5.3.11)) :

$$\sigma^2 \Big|_{\text{VC}} = \frac{\alpha(1 - \alpha)}{p^2(y_\alpha)} \times (1 - \rho_I^2)$$

et pour l'estimateur par stratification contrôlée adaptative à deux strates, avec la séparatrice des strates en  $\alpha$  :

$$\sigma^2 \Big|_{\text{SCA}} = \frac{\alpha(1 - \alpha)}{p^2(y_\alpha)} \times \left( 1 - \frac{\rho_I^2}{8\alpha(1 - \alpha)} + O(\rho_I^3) \right)$$

Ici,  $\rho_I$  est le coefficient de corrélation entre  $\mathbf{1}_{Y \leq y_\alpha}$  et  $\mathbf{1}_{Z \leq z_\alpha}$  donné par (5.3.12). Cela montre que la méthode par variable de contrôle et la méthode par stratification contrôlée adaptative apportent une réduction de variance du même ordre lorsqu'on cherche à estimer des quantiles

autour de la médiane  $\alpha \sim 1/2$ . Mais, lorsqu'on cherche à estimer des quantiles importants, pour lesquels  $\alpha \sim 0$  ou  $1$ , alors la méthode par stratification contrôlée adaptative apporte une bien meilleure réduction de variance.

#### 5.4.4.2 Simulations

##### Exemple 1 : fonction 1D

On reprend l'exemple 1 de la page 168 et on cherche le quantile à  $\alpha = 95\%$  de  $Y$ . On compare les performances de l'estimateur empirique, de l'estimateur avec variable de contrôle  $Z$ , et de l'estimateur par stratification contrôlée adaptatif. Pour ce dernier, on implémente d'abord la méthode avec deux strates  $[0, \alpha_1]$  et  $] \alpha_1, 1]$  avec une séparatrice en  $\alpha_1 = \alpha$ . On utilise  $\tilde{n} = 2 \times n/10$  simulations pour l'évaluation (5.4.17) de la répartition optimale, en mettant  $n/10$  simulations dans chacune des deux strates. Les résultats tirés d'une série de 10000 simulations peuvent être résumés de la manière suivante :

Estimation du $\alpha$ -quantile avec $\alpha = 0.95$ , $n = 2000$ , $y_\alpha = 3.66$			
estimation empirique :		moyenne	écart-type
	$\hat{Y}_{\alpha,n}$	3.6614	0.33473
estimation CV :		moyenne	écart-type
	$\hat{Y}_{\alpha,n}$	3.6502	0.29433
méthode SCA à 2 strates [0, 0.95], ]0.95, 1] :		moyenne	écart-type
	$\tilde{\beta}_1$	0.8562	0.01874
	$\hat{Y}_{\alpha,n}$	3.6495	0.27839

Remarquer que la méthode par stratification contrôlée adaptative décide ici d'allouer 85% des simulations dans la strate  $[0, 0.95[$ , et donc 15% dans la strate  $[0.95, 1]$ . On est donc bien ici dans un cadre où il faut augmenter le nombre de simulations dans la strate  $[0.95, 1]$  par rapport à ce qui est prévu dans le cas où il n'y a pas de contrôle, où on aurait seulement 5% de simulations dans la strate  $[0.95, 1]$ .

On a aussi appliqué la méthode de stratification contrôlée adaptative avec trois strates  $[0, \alpha_1[$ ,  $[\alpha_1, \alpha_2[$ ,  $[\alpha_2, 1]$ , avec des séparatrices en  $\alpha_1 = 0.85$  et  $\alpha_2 = 0.95$ . On utilise  $3 \times n/10$  simulations pour l'évaluation (5.4.17) de la répartition optimale, en mettant  $n/10$  simulations dans chacune des trois strates. Les résultats tirés d'une série de 10000 simulations peuvent être résumés de la manière suivante :

Estimation du $\alpha$ -quantile avec $\alpha = 0.95$ , $n = 2000$ , $y_\alpha = 3.66$			
		moyenne	écart-type
méthode SCA à 3 strates [0, 0.85], ]0.85, 0.95], ]0.95, 1]	:	$\tilde{\beta}_1$	0.10136
		$\tilde{\beta}_2$	0.57666
		$\tilde{\beta}_3$	0.32199
		$\hat{Y}_{\alpha,n}$	3.6509
		0.023284	0.016999
		0.0098153	0.11703

On peut ici noter une importante réduction de l'écart-type de l'estimateur par stratification contrôlée adaptative. On peut aussi remarquer que la répartition optimale devrait en fait avoir une fraction  $\beta_1 < 0.1$ , mais la borne 0.1 ne peut pas être franchie du fait qu'on a utilisé  $n/10$  simulations dans la strate  $[0, \alpha_1[$  durant la première phase de l'estimation.

Les simulations précédentes ont été réalisées avec des échantillons de taille  $n = 2000$ . Dans ce cas, la méthode SCA est très robuste, dans le sens où le choix du nombre  $\tilde{n}$  de simulations dédiées à l'estimation de l'allocation optimale n'est pas critique. Quand  $n$  est plus petit, comme par exemple  $n = 200$ , le choix de  $\tilde{n}$  devient plus critique : si  $\tilde{n}$  est trop petit, alors l'estimation de l'allocation optimale peut être mauvaise ; si  $\tilde{n}$  est trop grand, alors  $n - \tilde{n}$  peut être trop petit et il devient alors impossible d'allouer le nombre voulu de simulations dans chaque strate. Nous avons appliqué la méthode SCA avec  $n = 200$  avec l'exemple 1, et il est apparu que la méthode SCA avec  $\tilde{n} = n/10$  est encore très performante. Cependant, on ne peut pas jurer que ce sera le cas dans toutes les applications. Les résultats tirés d'une série de 10000 simulations sont les suivants :

Estimation du $\alpha$ -quantile avec $\alpha = 0.95$ , $n = 200$ , $y_\alpha = 3.66$			
estimation empirique :		moyenne	écart-type
		$\hat{Y}_{\alpha,n}$	3.8756
estimation VC :		moyenne	écart-type
		$\hat{Y}_{\alpha,n}$	3.7347
méthode SCA à 3 strates [0, 0.85], ]0.85, 0.95], ]0.95, 1]		moyenne	écart-type
	:	$\tilde{\beta}_1$	0.14331
		$\tilde{\beta}_2$	0.55261
		$\tilde{\beta}_3$	0.30408
		$\hat{Y}_{\alpha,n}$	3.6201
		0.16441	0.11466
		0.063331	0.37801

Les écart-types des estimations des  $\beta_j$  sont beaucoup plus grands ici que dans le cas  $n = 2000$ , mais la qualité de l'estimation de l'allocation optimale est encore assez bonne pour permettre une réduction de variance significative pour l'estimation du quantile. L'écart-type

de l'estimateur SCA est ici 2 fois plus petit comparé à l'estimateur empirique ou l'estimateur VC. Attention, si  $n = 100$ , alors la méthode SCA ne marche plus, car il commence à arriver des simulations où la méthode donne une mauvaise estimation de l'allocation optimale, et alors l'estimation du quantile peut être mauvaise. Dans ce cas, la méthode SC avec une allocation des niveaux prescrites par l'utilisateur est à recommander.

## 5.5 Estimation de quantile avec échantillonnage préférentiel

Nous allons reprendre les notations des sections précédentes.  $Y = f(X)$  est la variable de sortie d'un gros code numérique,  $X$  le vecteur aléatoire des paramètres d'entrée et  $Z = f_r(X)$  est un modèle réduit du code. Nous allons proposer un estimateur de la fonction de répartition par échantillonnage préférentiel. Cette méthode consiste à modifier la loi des variables aléatoires échantillonnées. La distribution de l'échantillon est alors concentrée aux endroits "d'importance".

Dans cette section, nous allons utiliser le modèle réduit pour trouver ces régions d'importance, et ainsi rechercher le  $\alpha$ -quantile du modèle complet en générant un vecteur aléatoire de taille réduite autour de ces régions.

### 5.5.1 Estimation de la fonction de répartition

#### 5.5.1.1 Estimation avec échantillonnage préférentiel

Un estimateur par échantillonnage préférentiel de la fonction de répartition est de la forme :

$$\hat{F}(y) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{Y_j \leq y\}} \frac{p(X_j)}{q(X_j)}, \quad X_j \sim q \quad (5.5.1)$$

où  $q$  est la densité d'importance.  $\hat{F}(y)$  converge presque sûrement vers  $F(y) = P(Y \leq y)$  quand  $n \rightarrow \infty$ . Pour que  $\hat{F}(y)$  soit un estimateur sans biais de la fonction de répartition, il faut que le support de la densité d'importance  $q$  contienne le support de la densité initiale  $p$ .

La variance de (5.5.1) est donnée par :

$$\text{Var}[\hat{F}(y)] = \frac{1}{n} \left( \int \frac{\mathbf{1}_{\{f(x) \leq y\}} p(x)^2}{q(x)} dx - F(y)^2 \right)$$

L'échantillonnage préférentiel ne garantit pas une réduction de variance par rapport à la simulation simple. Le choix de la densité d'importance  $q$  influe directement sur la variance de l'estimateur. En théorie, il est possible d'améliorer la précision de l'estimateur. Le minimum de la variance de  $\hat{F}(y)$  est atteint pour la densité d'importance suivante, dite densité optimale :

$$q^*(x) = \frac{\mathbf{1}_{\{f(x) \leq y\}} p(x)}{\int \mathbf{1}_{\{f(x') \leq y\}} p(x') dx'} \quad (5.5.2)$$

Ce résultat est inutilisable en pratique car le dénominateur de  $q^*$  est justement ce que nous cherchons à estimer. Pour le choix de la densité d'importance, il serait intéressant de s'approcher de la densité optimale.

### 5.5.1.2 Recherche de la densité d'importance

Comme nous l'avons montré dans le chapitre 2, pour que  $q$  soit une densité d'importance convenable, il faut que :

- $q$  “domine” la vraie distribution, c'est-à-dire que  $p > 0$  quand  $q > 0$  ;
- $X \sim q$  soit facile à générer ;
- $p(x)/q(x)$  soit facile à calculer ;
- $q$  minimise, autant que possible, la variance de l'estimateur.

Nous constatons qu'une manière commode de choisir la densité d'importance est de choisir tout d'abord une famille de densités  $\mathcal{G} = \{q_\gamma; \gamma \in \Gamma\}$  (approche paramétrique) et d'en adapter les paramètres. Cette famille de densités doit correspondre à des variables aléatoires faciles à générer. Dans la suite, nous supposons que  $X$  est une v.a. à valeurs dans  $\mathbb{R}^d$  et que la famille de densités  $\mathcal{G}$  est paramétrée par les deux premiers moments  $\gamma = (\lambda, C) : \lambda \in \mathbb{R}^d$  est l'espérance et  $C \in \mathcal{M}_d(\mathbb{R})$  est la matrice de covariance de  $X$  sous la loi de densité  $q_\gamma$ .

La stratégie pour déterminer la meilleure densité d'importance parmi la famille  $\mathcal{G}$  est la suivante. Nous connaissons la formule théorique (5.5.2) de la densité optimale  $q^*$ . La v.a.  $X$  sous la loi de densité  $q^*$  a pour espérance

$$\lambda^* = \frac{\int x \mathbf{1}_{\{f(x) \leq y\}} p(x) dx}{\int \mathbf{1}_{\{f(x) \leq y\}} p(x) dx} \quad (5.5.3)$$

et pour matrice de covariance

$$C^* = \frac{\int x x^t \mathbf{1}_{\{f(x) \leq y\}} p(x) dx}{\int \mathbf{1}_{\{f(x) \leq y\}} p(x) dx} - \lambda^* \lambda^{*t} \quad (5.5.4)$$

On choisit alors parmi la famille de densités  $\mathcal{G}$  la densité  $q_\gamma$  qui a pour moyenne  $\lambda^*$  et pour matrice de covariance  $C^*$ , c'est-à-dire la densité  $q_{\gamma^*}$  avec  $\gamma^* = (\lambda^*, C^*)$ .

Il reste donc à trouver un moyen d'estimer  $\lambda^*$  et  $C^*$ . Si nous avons à disposition un nombre quasi illimité de simulations utilisant le modèle réduit (supposé peu coûteux), nous pouvons estimer  $\lambda^*$  et  $C^*$  par :

$$\begin{cases} \hat{\lambda} = \frac{\sum_{i=1}^{\tilde{n}} X_i \mathbf{1}_{\{Z_i \leq y\}} p(X_i) / q_0(X_i)}{\sum_{i=1}^{\tilde{n}} \mathbf{1}_{\{Z_i \leq y\}} p(X_i) / q_0(X_i)} \\ \hat{C} = \frac{\sum_{i=1}^{\tilde{n}} X_i X_i^t \mathbf{1}_{\{Z_i \leq y\}} p(X_i) / q_0(X_i)}{\sum_{i=1}^{\tilde{n}} \mathbf{1}_{\{Z_i \leq y\}} p(X_i) / q_0(X_i)} - \hat{\lambda} \hat{\lambda}^t \end{cases}, \quad X_i \sim q_0 \quad (5.5.5)$$

où  $q_0$  est une densité choisie a priori dans la famille  $\mathcal{G}$  et  $\tilde{n} \gg 1$ . On estime en fait ici non pas  $\lambda^*$  et  $C^*$  mais

$$\lambda_r^* = \frac{\int x \mathbf{1}_{\{f_r(x) \leq y\}} p(x) dx}{\int \mathbf{1}_{\{f_r(x) \leq y\}} p(x) dx} \text{ et } C_r^* = \frac{\int x x^t \mathbf{1}_{\{f_r(x) \leq y\}} p(x) dx}{\int \mathbf{1}_{\{f_r(x) \leq y\}} p(x) dx} - \lambda_r^* \lambda_r^{*t}$$

que l'on suppose être proches de  $\lambda^*$  et  $C^*$  lorsque  $f_r$  est un modèle réduit raisonnable.

Enfin, il faut remarquer que la densité d'importance choisie dépend de  $y$ . On ne peut pas trouver une densité d'importance qui soit bonne universellement, i.e. pour toutes les valeurs de  $y$  en même temps. C'est tout-à-fait normal, car le principe de l'échantillonnage préférentiel est de favoriser les tirages qui tombent dans une région de l'espace des  $x$  qui soit importante pour la fonction cible dont on cherche à évaluer la moyenne (ici,  $x \mapsto \mathbf{1}_{f(x) \leq y}$ ).

### 5.5.2 Estimation du quantile à partir du modèle réduit

On cherche ici à estimer le  $\alpha$ -quantile de  $Y$ . Pour cela nous allons chercher une densité d'importance correcte pour le calcul de l'intégrale suivante :

$$I = \int_{\mathbb{R}^d} \mathbf{1}_{\{f_r(x) \leq z_\alpha\}} p(x) dx$$

où  $f_r$  est notre modèle réduit et  $z_\alpha$  est le  $\alpha$ -quantile de  $Z$ , supposé connu avec une précision arbitraire. Trouver la densité d'importance  $q$  qui minimise la variance de cet estimateur va nous permettre de nous approcher des régions d'importance de notre code numérique. En effet, si les  $\alpha$ -quantiles de  $f$  et de  $f_r$  ne sont pas trop éloignés, nous aurons des tirages dans la queue de distribution de  $Y$  autour du  $\alpha$ -quantile.

Comme précédemment, on recherche une bonne densité d'importance parmi une famille paramétrique  $\mathcal{G}$  de densités  $q_\gamma$  paramétrées par leurs deux premiers moments  $\gamma = (\lambda, C)$ . Si nous disposons d'un nombre quasi-illimité d'appels au modèle réduit, alors on utilise l'estimateur (5.5.5) avec  $y = z_\alpha$ . Puis nous utilisons l'estimateur par échantillonnage préférentiel (5.5.1) avec la densité d'importance  $q_{\hat{\gamma}}$ ,  $\hat{\gamma} = (\hat{\lambda}, \hat{C})$ . Enfin l'estimateur du  $\alpha$ -quantile est  $\hat{Y}_{\alpha, n} = \inf\{y, \hat{F}(y) \geq \alpha\}$ .

Si nous disposons seulement d'un nombre limité de simulations (modèle réduit un peu coûteux), nous pouvons utiliser le tirage d'importance adaptatif décrit dans le chapitre 2 : à chaque étape, le tirage d'importance adaptatif estime  $\gamma$  en prenant les évaluations obtenues par la méthode du tirage d'importance simple qui utilise des v.a. *i.i.d* suivant la nouvelle fonction d'importance ; et ce jusqu'à la précision souhaitée.

### 5.5.3 Simulations

#### Exemple 1 : fonction 1D

Soit  $X$  une v.a. gaussienne centrée réduite, les fonctions  $f$  et  $f_r$  sont données par :

$$f_r(x) = |x|x, \quad f(x) = 0.95|x|x[1 + 0.5 \cos(10x) + 0.5 \cos(20x)] \quad (5.5.6)$$

Ces fonctions et leur densité de probabilité sont illustrées en figure 5.5.1 (a, b). On souhaite déterminer le  $\alpha$ -quantile de  $Y$ . On dispose d'un nombre limité  $n$  d'appels à la fonction  $f$  et d'un nombre quasi-illimité d'appels à la fonction  $f_r$ . On utilise la méthode d'échantillonnage préférentiel paramétrique pour déterminer la densité d'importance  $q$  à partir du modèle réduit, et on l'utilise pour le calcul du  $\alpha$ -quantile de  $Y$ . On compare le quantile obtenu avec le quantile empirique en figure 5.5.1c, et on constate qu'il y a une réduction de la variance.

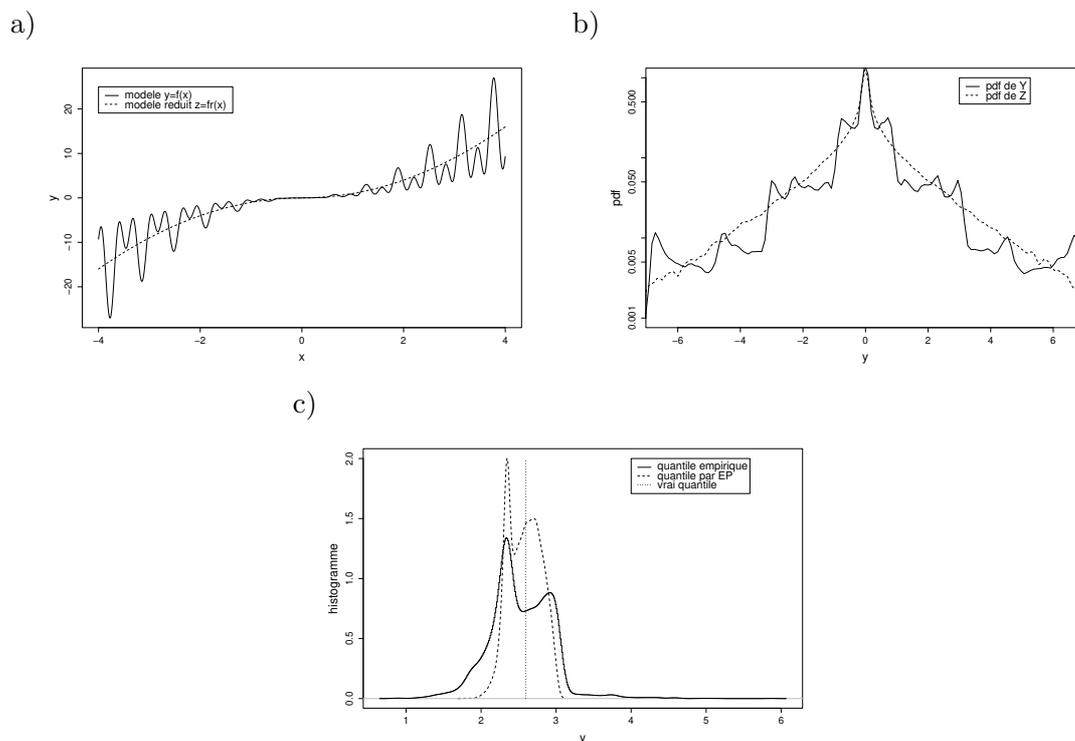


FIG. 5.5.1 – Figure a : graphes des fonctions  $f$  et  $f_r$  données par (5.5.6). Figure b : pdf de  $Y = f(X)$  et  $Z = f_r(X)$  pour  $X \sim \mathcal{N}(0,1)$ . Figure c : estimations du  $\alpha$ -quantile pour  $Y$ ,  $\alpha = 0.95$  à partir d'un échantillon de taille  $n = 200$ . Les histogrammes sont tracés en utilisant les résultats d'une série de 5000 expériences.

### Exemple 2 : fonction 2D

Soit  $X = (X_1, X_2)$  un couple de v.a. gaussiennes centrées réduites indépendantes, les fonc-

tions  $f$  et  $f_r$  sont données par :

$$f_r(x) = |x_1|x_1 + x_2 \quad (5.5.7)$$

$$f(x) = 0.95|x_1|x_1[1 + 0.5 \cos(10x_1) + 0.5 \cos(20x_1)] \\ + 0.7x_2[1 + 0.4 \cos(x_2) + 0.3 \cos(14x_2)] \quad (5.5.8)$$

Ces fonctions et leur densité de probabilité sont illustrées en figure 5.5.2 (a, b, c). On compare le quantile obtenu par échantillonnage préférentiel avec le quantile empirique en figure 5.5.2d, et on constate une nette réduction de la variance.

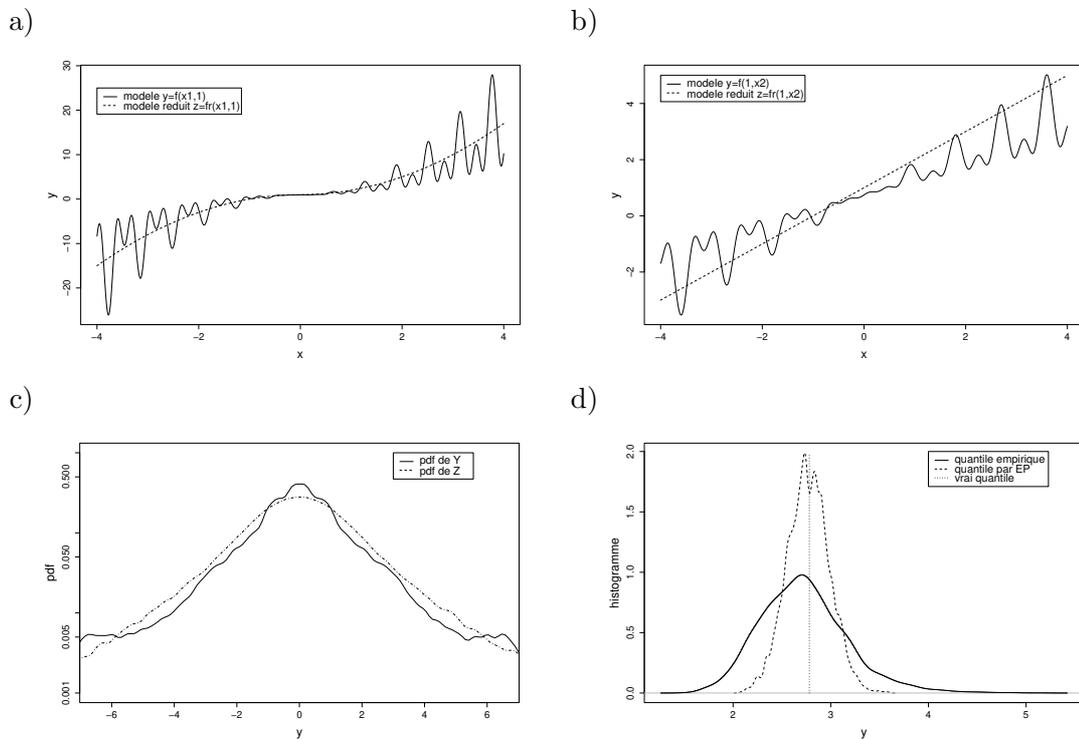


FIG. 5.5.2 – Figures a,b : graphes en coupe des fonctions  $f(x_1, x_2)$  et  $f_r(x_1, x_2)$  données par (5.5.8-5.5.7). Figure c : pdf de  $Y = f(X_1, X_2)$  et  $Z = f_r(X_1, X_2)$  pour  $X_1, X_2 \sim \mathcal{N}(0, 1)$ . Figure d : estimation du  $\alpha$ -quantile pour  $Y$ ,  $\alpha = 0.95$  à partir d'un échantillon de taille  $n = 200$ . Les histogrammes sont tracés pour 5000 expériences.

## 5.6 Implémentation avec le code CATHARE

### 5.6.1 Présentation du code

Dans notre cadre, le code CATHARE du CEA est utilisé pour calculer le premier pic de température de la gaine du combustible d'un REP (Réacteur à Eau sous Pression), lors d'un

scénario d'accident APRP - GB (Accident de Perte de Réfrigérant Primaire - Grosse Brèche).

On a donc affaire à une fonction  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  où

- $\mathbb{R}^p$  représente l'espace des paramètres d'entrée,
- $\mathbb{R}$  représente l'espace de la réponse du code (i.e. la variable de sortie qui est ici le pic de température),
- $f$  est la boîte noire qui représente le code ; son temps de calcul est d'environ une vingtaine de minutes sur une station de travail.

Certains paramètres d'entrée du code sont incertains, on les modélise par un vecteur aléatoire  $X = (X_1, \dots, X_p)$ . Il existe  $p = 53$  paramètres incertains, décrits par des v.a. réelles  $X_i$  indépendantes et de lois diverses (normales et log-normales). Le scénario étudié a été défini et mise en œuvre par Bazin and de Crecy [BdC 06] lors de l'exercice international BEMUSE (projet de l'Agence de l'Energie Nucléaire de l'OCDE).

### 5.6.2 Détermination d'une surface de réponse

La surface de réponse que l'on doit ajuster doit satisfaire aux trois contraintes suivantes :

- être suffisamment corrélée avec le modèle,
- avoir un temps de calcul négligeable (temps pour une évaluation de la surface de réponse),
- être construite avec un minimum de calculs (moins de 50 pour pouvoir ensuite appliquer la stratification contrôlée avec 200 calculs).

Avant de chercher à ajuster une surface de réponse sur CATHARE, et vu le grand nombre de paramètres d'entrée aléatoire, il est nécessaire de déterminer quels sont les paramètres les plus influents (étape dite de "screening").

#### 5.6.2.1 Screening

Pour réaliser ce screening, Roger Phan-Tan-Luu (Université Paul Cézanne, Marseille) nous a proposé d'utiliser un plan d'expériences par matrice supersaturée [Sat 00]. Cette méthode est la moins coûteuse quand on ne sait rien sur le sens de variation de la sortie par rapport aux entrées du modèle. Le nombre d'expériences à réaliser est inférieur au nombre de facteurs (les paramètres d'entrée). Elle suppose cependant que le nombre de facteurs réellement influents est faible devant le nombre de facteurs étudiés, et qu'il n'y a pas d'effet non linéaire ni d'effet dû à des interactions entre facteurs. Elle suppose également que les facteurs ont deux niveaux : un niveau bas et un niveau haut. Etant donné qu'ici chacun des facteurs est une v.a. représentée par une densité de probabilité, on choisit pour les niveaux bas et haut les quantiles à 25% et à 75%.

La matrice supersaturée est construite par la méthode de Lin [Lin 93] améliorée par R. Phan-Tan-Luu et M. Sergent (publication en cours de rédaction). Pour 53 facteurs, ce plan d'expériences nécessite 30 calculs. Lors de l'exécution de ces 30 calculs, on constate que 5 calculs ne donnent pas de résultat du fait de problème de convergence du code CATHARE. Après analyse de la matrice d'expériences des calculs qui ont échoué, et quelques tests supplémentaires, le facteur problématique a pu être isolé (paramètre n° 36, "nucleated boiling - Heat Transfer Coefficient"). En remplaçant ce facteur par une valeur tronquée, on obtient des résultats pour les calculs problématiques. Un pari a lieu ici : il faut que ce paramètre d'entrée n'ait pas d'influence sur la sortie, ce qui semble vrai au vu des premiers résultats et des avis des experts. Dans les simulations par stratification contrôlée, sa valeur sera tronquée afin que les calculs n'échouent plus.

L'analyse des résultats de la matrice supersaturée a permis de mettre à jour 5 paramètres probablement importants et de donner l'ordre de grandeur de l'effet (ainsi que son signe) :

n° paramètre	19	44	9	42	20
Effet	-53	-48	32	-31	28

### 5.6.2.2 Régression

Usuellement, la détermination d'une surface de réponse ne se fait pas avec les calculs utilisés pour le screening, mais avec un plan d'expériences construit spécifiquement pour le type de surfaces de réponse que l'on veut ajuster. Cependant, vu la limitation en nombre de calculs que nous impose notre étude, nous choisissons de passer outre la rigueur et d'utiliser les calculs du plan supersaturé pour déterminer une surface de réponse. La construction rigoureuse d'une surface de réponse sera détaillée dans une prochaine étude.

Sur la base des 30 calculs, nous effectuons donc une régression linéaire de la sortie sur les 5 facteurs influents :

```
-----
| Les coefficients de la regression |
-----
```

	Estimation	Ecart-type	t-value	Pr(> t )
Intercept	5.4288e+02	2.0165e+02	2.6923e+00	1.2730e-02
V 9	7.4480e+02	1.6570e+02	4.4948e+00	1.5009e-04

V 19	-4.6120e+02	6.1681e+01	-7.4771e+00	1.0220e-07
V 20	2.5010e+02	6.1616e+01	4.0590e+00	4.5391e-04
V 42	-2.1483e+02	5.1316e+01	-4.1864e+00	3.2864e-04
V 44	-3.5101e+01	5.2659e+00	-6.6658e+00	6.7893e-07

Le coefficient de détermination de cette régression vaut  $R^2 = 0.87$ . Pour connaître les capacités de prédiction de la surface de réponse, on simule 1000 répliques bootstrap de la base initiale (celle de 30 calculs) et on effectue la régression sur chacune d'elle. Le  $R^2$  en prédiction est estimé à partir des calculs non utilisés dans chacune des régressions :

$$R_{pred}^2 = 0.72, \text{ avec un intervalle de confiance à 90\% de } [0.43, 0.88].$$

Il est probable que certains facteurs, qui ont une influence moyenne, ont été omis par l'analyse de la matrice supersaturée. Pour les identifier, on effectue alors une régression pas-à-pas ("stepwise regression"), de manière manuelle, en ajoutant progressivement d'autres facteurs aux 5 premiers facteurs. Au final, on parcourt les 48 facteurs restant. On décide de garder ou de rejeter chaque facteur en analysant sa t-value de la régression linéaire. On conserve au final 10 facteurs et on obtient :

```

-----
| Les coefficients de la regression |
-----

```

	Estimation	Ecart-type	t-value	Pr(> t )
Intercept	6.6030e+02	1.4979e+02	4.4081e+00	3.0210e-04
V 2	-6.1792e+01	3.0437e+01	-2.0302e+00	5.6576e-02
V 6	6.1414e+00	2.3431e+00	2.6210e+00	1.6813e-02
V 9	5.8996e+02	1.3108e+02	4.5009e+00	2.4468e-04
V 11	8.0820e+01	3.8636e+01	2.0918e+00	5.0120e-02

V 19	-4.0455e+02	4.5450e+01	-8.9010e+00	3.3159e-08
V 20	2.6423e+02	4.5266e+01	5.8373e+00	1.2703e-05
V 35	-2.7056e+01	1.4173e+01	-1.9090e+00	7.1489e-02
V 37	6.1613e+00	3.3215e+00	1.8550e+00	7.9185e-02
V 42	-2.5572e+02	3.6122e+01	-7.0793e+00	9.7998e-07
V 44	-3.1987e+01	4.0197e+00	-7.9576e+00	1.8135e-07

Le coefficient de détermination de cette régression vaut  $R^2 = 0.95$ . Pour connaître les capacités de prédiction de la surface de réponse, on simule 1000 répliques bootstrap de la base initiale (celle de 30 calculs) et on effectue la régression sur chacune d'elle. Le  $R^2$  en prédiction est estimé à partir des calculs non utilisés dans chacune des régressions :

$$R_{pred}^2 = 0.77, \text{ avec un intervalle de confiance à 90\% de } [0.46, 0.92].$$

L'augmentation de 5% du  $R_{pred}^2$  entre la surface de réponse à 10 termes et celle à 5 termes n'a donc pas été très spectaculaire, mais elle nous semble suffisamment significative pour que l'on utilise le modèle à 10 termes dans la stratification contrôlée.

Le bilan de cette étude est qu'un modèle réduit  $f_r$  a pu être obtenu à l'issue d'une série de 30 calculs. C'est une forme linéaire en les 10 paramètres importants déterminés ci-dessus :

$$Z = f_r(X)$$

avec

$$\begin{aligned} f_r(X) = & 660.30 - 61.792X_2 + 6.1414X_6 + 589.96X_9 + 80.82X_{11} - 404.55X_{19} \\ & + 264.23X_{20} - 27.056X_{35} + 6.1613X_{37} - 255.72X_{42} - 31.987X_{44} \end{aligned} \quad (5.6.1)$$

Ce modèle est raisonnable dans le régime normal de fonctionnement.

### 5.6.2.3 Analyse de la validité de la surface de réponse

Afin de quantifier de manière précise la qualité des estimateurs des quantiles dans la section suivante, une base de 1000 simulations a été produite : génération aléatoire de 1000 jeux de paramètres d'entrée (chaque paramètre suivant sa loi de probabilité), puis réalisation des calculs

CATHARE afin d'obtenir un échantillon de 1000 réponses. Il est clair que dans les applications pratiques de notre méthodologie, de telles bases de validation ne seront pas faisables.

Dans ce paragraphe, on va utiliser cette base de 1000 calculs pour quantifier de manière plus précise la qualité de la surface de réponse obtenue à partir uniquement de 30 calculs. Sur cette base de 1000 calculs, on a 1000 réponses du code et 1000 prédictions faites par la surface de réponse. La figure 5.6.1a permet de visualiser l'adéquation entre le code et la surface de réponse. On constate également qu'il y a quelques points très mal prédits. La figure 5.6.1b confirme la qualité de la régression sur la base des points de la matrice supersaturée.

On peut également calculer le coefficient de corrélation entre le code et la surface de réponse :

$$\rho = 0.60 \text{ pour la surface de réponse à cinq facteurs,}$$

$$\rho = 0.66 \text{ pour la surface de réponse à dix facteurs (équation (5.6.1)).}$$

Comme on l'avait observé sur la base des calculs de la matrice supersaturée, la surface de réponse à 10 termes représente légèrement mieux le code que celle à 5 termes.

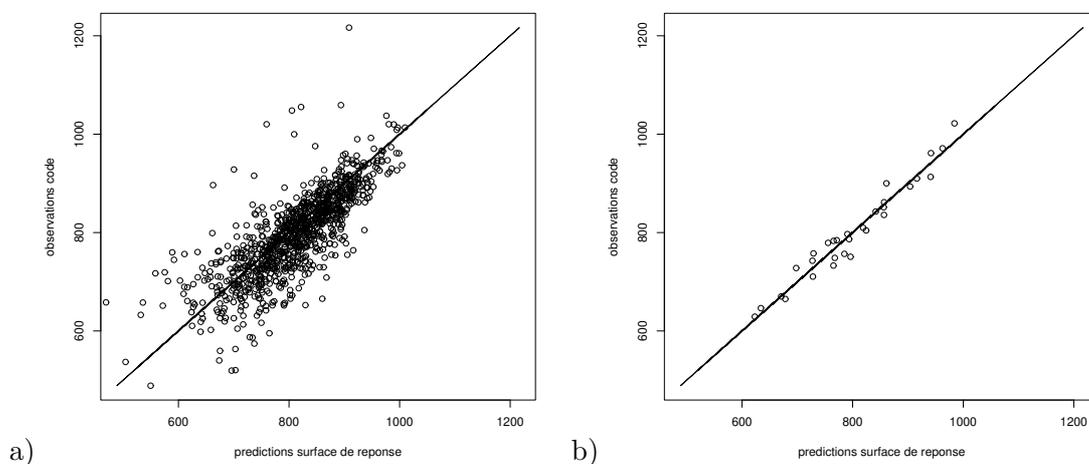


FIG. 5.6.1 – Comparaisons entre les valeurs obtenues par le code et les valeurs prédites par la surface de réponse (5.6.1). Figure a : base de test (les 1000 calculs supplémentaires). Figure b : base de construction (les calculs de la matrice supersaturée).

Nous avons également testé la construction de surfaces de réponse par une méthodologie basée sur les plans d'expériences classiques [DFS 97], en nous limitant aux 5 facteurs les plus influents (ceux déterminées par la matrice supersaturée). Les 48 autres facteurs sont fixés à leur valeur nominale. R. Phan-Tan-Luu nous a fourni une matrice d'expériences de type hybride dans un domaine sphérique (29 calculs supplémentaires) et une matrice d'expériences

composite dans un domaine cubique (27 calculs supplémentaires). Pour chacun de ces plans, on obtient une surface de réponse polynomiale d'ordre deux à 21 termes (tous les termes d'ordre un et deux sont présents). On teste ces surfaces de réponse sur la base des 1000 calculs. Le coefficient de corrélation entre le code et la surface de réponse vaut :

$$\begin{aligned}\rho &= 0.50 \text{ pour la surface de réponse définie dans un domaine sphérique,} \\ \rho &= 0.13 \text{ pour la surface de réponse définie dans un domaine cubique.}\end{aligned}$$

Le principe de ces techniques est de bien représenter le phénomène à l'intérieur du domaine que l'on a défini par des niveaux bas et haut sur chaque facteur. Il est donc formellement déconseillé de les utiliser en extrapolation. Or, les niveaux bas et haut sont ici les quantiles à 5% et à 95% (on pourrait prendre des quantiles plus extrêmes, mais cela pose alors le problème d'intervalles de variation par facteur beaucoup trop grands). Notre base de 1000 points contient donc environ 10% de points à l'extérieur de ces quantiles, pour lesquels les surfaces de réponse sont utilisées en extrapolation. Ceci explique la dégradation du coefficient de corrélation par rapport au modèle plus rudimentaire (mais plus robuste) d'ordre un (équation (5.6.1)) déterminée par régression linéaire. Ce type de plans pour construire des surfaces de réponse ne semble donc pas bien adapté à notre problème qui consiste à chercher des valeurs dans les queues de distribution des paramètres d'entrée.

### 5.6.3 Stratification contrôlée

#### 5.6.3.1 Stratégie à quatre strates

On a utilisé une stratégie à quatre strates :

$$\alpha = 0.95 \longleftrightarrow z_\alpha = 936.05^\circ\text{C}$$

$$\alpha = 0.85 \longleftrightarrow z_\alpha = 890.30^\circ\text{C}$$

$$\alpha = 0.55 \longleftrightarrow z_\alpha = 821.15^\circ\text{C}$$

Les quantiles de  $Z$  ont été obtenus par simulation de Monte Carlo avec  $10^7$  échantillons de  $X$ .

#### Premier essai :

Le premier essai a donné les résultats suivants, pour  $\alpha = 0.95$  et  $n = 200$  (figure 5.6.2 a)) :

$$\text{Quantile estimé : } \quad 928^\circ\text{C}$$

$$\text{Ecart-type de l'estimation (par bootstrap) : } \quad 7^\circ\text{C}$$

#### Deuxième essai :

On a refait cette expérience dans les mêmes conditions, et on a obtenu, toujours pour  $\alpha = 0.95$  et  $n = 200$  (figure 5.6.2 b)) :

Quantile estimé :  $934^{\circ}\text{C}$

Ecart-type de l'estimation (par bootstrap) :  $11^{\circ}\text{C}$

### Troisième essai :

Enfin, on a regroupé nos deux jeux de simulations, ce qui donne une expérience avec 400 simulations. Cet essai a donné les résultats suivants, pour  $\alpha = 0.95$  et  $n = 400$  (figure 5.6.2 c)) :

Quantile estimé :  $930^{\circ}\text{C}$

Ecart-type de l'estimation (par bootstrap) :  $6.5^{\circ}\text{C}$

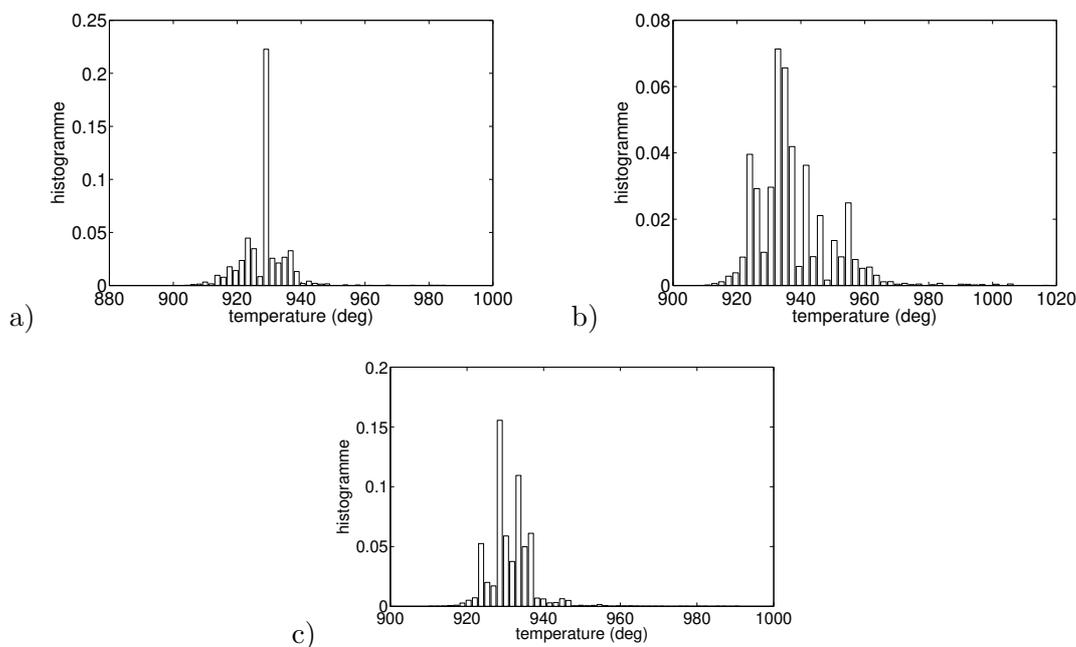


FIG. 5.6.2 – Histogramme du quantile estimé obtenu par bootstrap sur 50000 ré-échantillonnages. Figure a) premier essai ; figure b) deuxième essai, c) troisième essai.

### 5.6.3.2 Stratégie à deux strates

On a essayé une stratégie à deux strates :

$$\alpha = 0.9 \longleftrightarrow z_{\alpha} = 908.70^{\circ}\text{C}$$

Une stratégie à deux strates fait peu confiance au modèle réduit pour le contrôle. On ne peut donc pas espérer une grosse réduction de variance par rapport au quantile empirique usuel. Cet essai a donné les résultats suivants (figure 5.6.3.2) :

Quantile estimé : 929°C  
 Ecart-type de l'estimation (par bootstrap) : 28°C

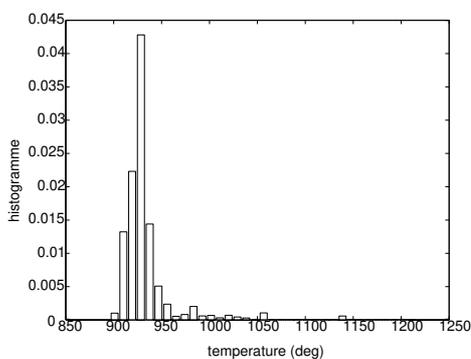


FIG. 5.6.3 – Histogramme du quantile empirique obtenu par bootstrap sur 50000 ré-échantillonnages.

### 5.6.3.3 Stratégie à cinq strates

On a essayé une stratégie à cinq strates pour déterminer le 0.95-quantile de  $Y$  avec  $n = 200$  échantillons :

$$\alpha = 0.95 \longleftrightarrow z_{\alpha} = 936.05^{\circ}\text{C}$$

$$\alpha = 0.9 \longleftrightarrow z_{\alpha} = 908.70^{\circ}\text{C}$$

$$\alpha = 0.8 \longleftrightarrow z_{\alpha} = 875.70^{\circ}\text{C}$$

$$\alpha = 0.6 \longleftrightarrow z_{\alpha} = 831.10^{\circ}\text{C}$$

Cette stratégie à cinq strates fait confiance au modèle réduit pour le contrôle. Or, ici, le maximum absolu 1078°C sur les 200 échantillons a été obtenu dans la dernière strate  $Z < z_{0.6}$ . Un autre pic très élevé, 978°C, a été obtenu dans cette strate. C'est pourquoi l'écart-type est grand, comme on le voit sur les résultats obtenus (figure 5.6.4) :

Quantile estimé : 947°C  
 Ecart-type de l'estimation (par bootstrap) : 28°C

On peut donc en conclure que, vu le niveau de corrélation du modèle réduit avec le modèle original, la stratégie à quatre strates apparaît ici comme étant la meilleure.

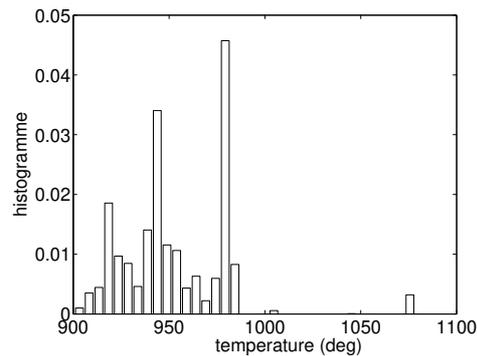


FIG. 5.6.4 – *Histogramme du quantile empirique obtenu par bootstrap sur 50000 ré-échantillonnages.*

#### 5.6.4 Echantillonnage préférentiel

On utilise une densité d'importance dont seuls les moments des lois des paramètres influents ont été modifiés (les paramètres de la surface de réponse) :

$$x_2 \sim \mathcal{N}(0.95, 0.15)$$

$$x_6 \sim \mathcal{N}(0.5, 2)$$

$$x_9 \sim \mathcal{N}(1.025, 0.038)$$

$$x_{11} \sim \mathcal{N}(1.05, 0.13)$$

$$x_{19} \sim \mathcal{N}(0.95, 0.1)$$

$$x_{20} \sim \mathcal{N}(1.08, 0.1)$$

$$x_{35} \sim \mathcal{LN}(-0.06, 0.31)$$

$$x_{37} \sim \mathcal{LN}(0.4, 1.15)$$

$$x_{42} \sim \mathcal{N}(0.95, 0.125)$$

$$x_{44} \sim \mathcal{LN}(-0.5, 0.93)$$

Ces valeurs ont été obtenues à partir de la formule (5.5.5). Le  $\alpha$ -quantile de  $Z$  a été obtenu avec  $10^6$  simulations et les nouveaux paramètres des lois ont été calculés avec un échantillon de taille 5000.

Pour les loi Log-normales, nous avons utilisé les relations suivantes, avec  $X \sim \mathcal{LN}(\mu, \sigma)$  :

$$\sigma^2 = \log \left( \frac{\mathbb{E}(X^2) + \mathbb{E}(X)^2}{\mathbb{E}(X)^2} \right)$$

$$\mu = \log(\mathbb{E}(X)) - 0.5\sigma^2$$

**Premier essai :**

Le premier essai a donné les résultats suivants, pour  $\alpha = 0.95$  et  $n = 200$  (figure 5.6.5 a)) :

Quantile estimé : 929°C  
 Ecart-type de l'estimation (par bootstrap) : 10.3°C

**Deuxième essai :**

On a refait cette expérience dans les mêmes conditions, et on a obtenu, toujours pour  $\alpha = 0.95$  et  $n = 200$  (figure 5.6.5 b)) :

Quantile estimé : 924°C  
 Ecart-type de l'estimation (par bootstrap) : 7.8°C

**Troisième essai :**

Enfin, on a regroupé nos deux jeux de simulations, ce qui donne une expérience avec 400 simulations. Cet essai a donné les résultats suivants, pour  $\alpha = 0.95$  et  $n = 400$  (figure 5.6.5 c)) :

Quantile estimé : 926.3°C  
 Ecart-type de l'estimation (par bootstrap) : 6.2°C

La densité d'importance choisie semble correcte. Elle a été « réglée » sur l'estimation du quantile du modèle réduit. Pour obtenir des résultats acceptables sur le modèle réduit, il a fallu contrôler le rapport de vraisemblance, c'est-à-dire, trouver un échantillon qui satisfaisait les conditions suivants :

$$\left| \sum_{i=1}^n p(X_i)/q(X_i) - n \right| \leq n\epsilon$$

$$\frac{\max(p(X_i)/q(X_i))}{\sum_{i=1}^n p(X_i)/q(X_i)} \leq \eta$$

Dans les cas précédents,  $\epsilon = 0.03$  et  $\eta = 0.08$

**5.6.5 Comparaisons à un estimateur empirique**

On compare les résultats de la stratification contrôlée et de l'échantillonnage préférentiel sur la base des 1000 calculs déjà utilisés pour examiner la qualité de la surface de réponse. On a appliqué une stratégie de Monte Carlo classique avec estimateur empirique pour déterminer le 0.95-quantile de  $Y$  avec  $n = 1000$  échantillons. On a obtenu :

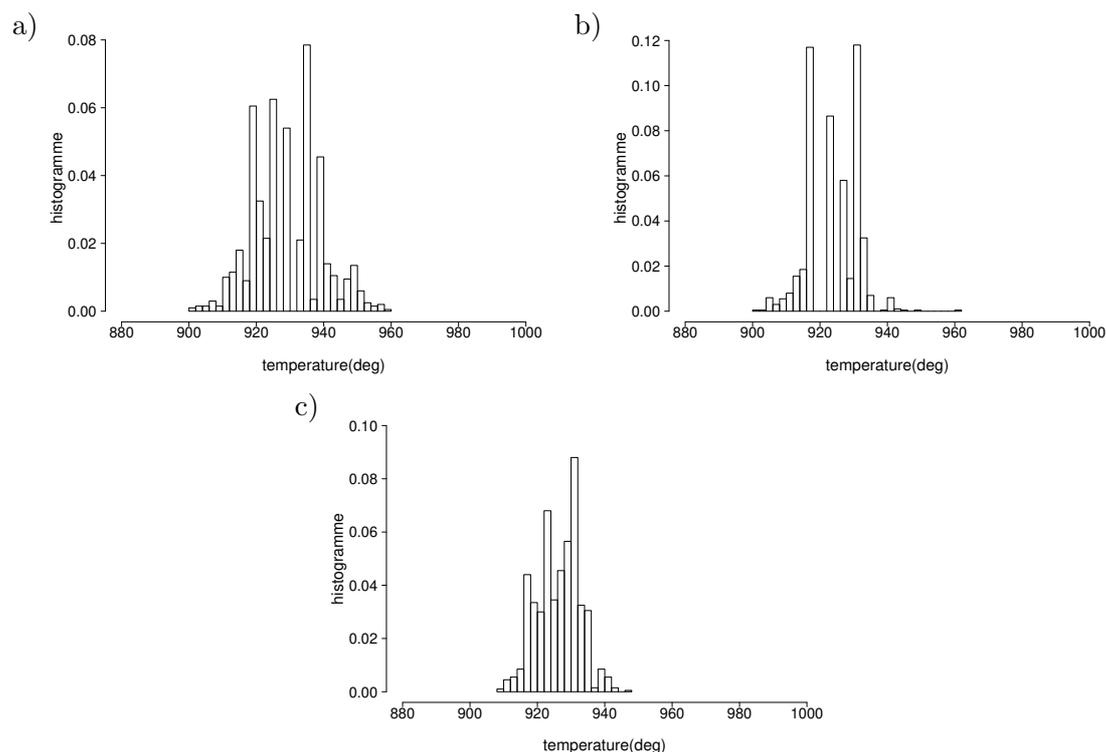


FIG. 5.6.5 – *Histogramme du quantile estimé obtenu par bootstrap sur 10000 ré-échantillonnages. Figure a) premier essai ; figure b) deuxième essai, c) troisième essai.*

Quantile estimé :  $928^{\circ}\text{C}$   
 Ecart-type de l'estimation (par bootstrap) :  $6.3^{\circ}\text{C}$

Le quantile estimé par l'estimateur empirique avec un échantillon de 1000 tirages donne donc une valeur en accord avec celles prédites par la méthode par stratification contrôlée (à deux ou quatre strates) et par la méthode par échantillonnage préférentiel. D'autre part, l'écart-type obtenu par bootstrap sur les 1000 échantillons classiques est du même ordre de grandeur (un tout petit peu plus grand), que l'écart-type obtenu par bootstrap sur les 400 échantillons stratifiés (§5.6.3.1, troisième essai), ainsi que du même ordre de grandeur que l'écart-type obtenu par bootstrap sur les 400 échantillons de l'échantillonnage préférentiel.

En utilisant les 1000 calculs de validation, on souhaite à présent regarder comment se comporte l'écart-type du quantile en fonction de la taille de l'échantillon (variant de 100 à 1000 par pas de 100). Pour une taille d'échantillon  $n$  fixée, on calcule moyenne et écart-type sur 10000 répliques. Chaque réplique consiste en  $n$  tirages avec remise dans la base des 1000 échantillons. Le tableau 5.6.1 fournit les résultats.

En comptant le coût de la construction de la surface de réponse, nos stratégies de stratification contrôlée et d'échantillonnage préférentiel nous ont coûté  $n = 230$  calculs. On peut donc

n	100	200	300	400	500	600	700	800	900	1000
Moyenne (en °C)	924	926	927	927	928	928	928	928	928	928
Ecart-type (en °C)	16.7	12.3	10.3	9.2	8.4	7.6	7.2	6.8	6.5	6.3

TAB. 5.6.1 – Estimation du quantile empirique et de son écart-type en fonction de la taille d'échantillons.

comparer ces résultats à l'estimateur empirique avec  $n = 200$ , qui nous offre une estimation du quantile à  $926^{\circ}\text{C}$  (léger biais par rapport à la valeur de référence  $928^{\circ}\text{C}$ ), avec un écart-type d'environ  $12.3^{\circ}\text{C}$ . La premier essai de la stratification contrôlée ne possède pas ce biais ( $928^{\circ}\text{C}$ ) et donne un écart-type presque deux fois plus petit ( $7^{\circ}\text{C}$ ); par contre, le deuxième essai est plus biaisé ( $934^{\circ}\text{C}$ ) et donne un écart-type du même ordre de grandeur ( $11^{\circ}\text{C}$ ). La premier essai de l'échantillonnage préférentiel possède un léger biais ( $929^{\circ}\text{C}$ ) et donne un écart-type du même ordre de grandeur ( $10.3^{\circ}\text{C}$ ); en revanche le deuxième essai est plus biaisé ( $924^{\circ}\text{C}$ ) et donne un écart-type bien plus petit ( $7.8^{\circ}\text{C}$ ). Le troisième essai peut être comparé à l'estimateur empirique avec  $n = 400$ , qui nous offre une estimation du quantile à  $927^{\circ}\text{C}$  avec un écart-type d'environ  $9.2^{\circ}\text{C}$ . Sur cet essai, l'écart-type de la stratification contrôlée est presque deux fois plus petit ( $5.5^{\circ}\text{C}$ ); l'estimation par échantillonnage préférentiel possède un léger biais ( $926.3^{\circ}\text{C}$ ) et donne un écart-type bien plus petit ( $6.2^{\circ}\text{C}$ ).

Finalement, sur cette étude où la surface de réponse est assez grossière (coefficient de corrélation  $\rho = 0.65$ ), on voit que la stratification contrôlée (à quatre strates) est capable de prédire le quantile sans biais, et de réduire l'écart-type de son estimation d'un facteur deux. On n'a pas implémenté la méthode SCA, car sa mise au point est venue trop tardivement par rapport au travail sur Cathare. Cependant, il serait sûrement très intéressant d'implémenter la méthode SCA à trois strates. On remarque enfin que l'échantillonnage préférentiel peut donner de bons résultats (un peu moins bons que la stratification contrôlée), à condition de choisir et de régler convenablement la densité d'importance et de contrôler le rapport de vraisemblance. Sans cela, cette méthode peut donner des résultats non présentables.

## 5.7 Conclusion

Dans ce chapitre, un panorama des méthodes usuelles d'estimations de quantiles a été donné. Les quantiles que nous avons envisagés ne sont pas des quantiles extrêmes (se mesurant avec des puissances de 10), mais plutôt des quantiles de l'ordre de 95%, avec des échantillons de taille de l'ordre de 200. On recherche par contre une bonne précision dans l'estimation du

quantile, et/ou un certain niveau de sureté. Les méthodes adaptées aux quantiles extrêmes, du type estimateur de Hill, n'ont donc pas été discutées ; l'étude s'est concentrée sur le quantile empirique, le quantile de Wilks, et l'intervalle de confiance du quantile empirique obtenu par bootstrap.

On a ensuite discuté différentes méthodes de réduction de variance dans le cas où :

- la variable d'intérêt  $Y = f(X)$  est la sortie d'un gros code numérique comportant un assez grand nombre de paramètres incertains  $X$ ,
- un modèle réduit  $Z = f_r(X)$  bien maîtrisé est disponible. On suppose qu'on connaît les quantiles de  $f_r(X)$  et qu'on peut appeler la fonction  $f_r$  un grand nombre de fois si on le souhaite.

On a d'abord examiné des méthodes de réduction de variance basées sur l'utilisation de  $Z$  (ou d'une fonction de  $Z$ ) comme variable de contrôle. On a montré que toutes ces méthodes permettent une réduction de la variance de l'estimateur du quantile en affectant aux  $n$  simulations  $Y_i = f(X_i)$ ,  $i = 1, \dots, n$  des poids différents selon les valeurs de  $Z_i = f_r(X_i)$ . Dans le cadre de ces méthodes, on effectue  $n$  tirages  $X_i$ ,  $i = 1, \dots, n$ , et on calcule les valeurs  $f(X_i)$  et  $f_r(X_i)$ .

On a ensuite examiné une méthode originale de stratification contrôlée. Dans ce cas, on tire par rejet les  $X_i$ , le rejet étant basé sur les valeurs  $f_r(X_i)$ . On fait donc beaucoup plus de simulations de variables  $X$  et d'appels au modèle réduit  $f_r$  que dans le cas des méthodes par variables de contrôle. Dans le cas de la stratification contrôlée, on contrôle exactement le nombre de réalisations de  $X_i$  qui tombent dans des régions bien délimitées par le modèle réduit, et on est ramené à estimer des probabilités conditionnelles. La réduction de variance sur l'estimateur du quantile peut alors être substantielle. Nous avons développé une méthode de stratification contrôlée adaptative dans laquelle les réalisations des paramètres d'entrée sont échantillonnées dans des zones déterminées par le modèle réduit, et dans laquelle le nombre de réalisations allouées dans chaque zone est optimisé de manière adaptative. La réduction de variance est alors très substantielle. Nous avons montré, par une analyse asymptotique des variances des estimateurs d'une part et par des simulations d'autre part, que la méthode SCA est la plus efficace dans les configurations que nous avons envisagées.

Enfin, on a examiné une méthode d'échantillonnage préférentiel. On tire les  $X_i$  suivant une nouvelle densité  $q$  convenablement choisie à partir du modèle réduit ; ensuite, on tire un échantillon  $X \sim q$  dont le rapport de vraisemblance est satisfaisant, la somme des rapport de vraisemblance doit être de l'ordre de la taille de l'échantillon et le maximum du rapport de vraisemblance ne doit pas dépasser un certain seuil. La réduction de variance sur l'estimateur

du quantile peut alors être relativement importante.

Les méthodes de stratification contrôlée et d'échantillonnage préférentiel ont été appliquées sur un problème fourni par le CEA (Cadarache et Grenoble) : l'estimation du quantile à 95% du premier pic de température de gaine lors d'un scénario accidentel dans le coeur d'une centrale nucléaire (APRP - GB). La coût en temps de calcul d'une évaluation du code et le grand nombre de paramètres d'entrée considérés nous ont amené à effectuer une analyse de sensibilité préliminaire à l'aide d'une matrice supersaturée. Il s'agit d'une méthode bien connue dans le domaine des plans d'expériences classiques, mais originale pour un plan d'expériences numériques. On a pu ensuite construire une surface de réponse linéaire d'ordre un représentant approximativement le code de calcul.

Pour notre test ( $n = 200$  et  $\alpha = 0.95$ ), l'écart-type de l'estimateur empirique peut être réduit d'un facteur deux par l'estimateur de la stratification contrôlée. Il serait utile d'avoir des critères quantitatifs pour choisir le nombre de strates optimal en fonction de la qualité de la surface de réponse. D'autre part, sur le même type de scénario (APRP - GB), Ardillon et de Rocquigny [AdR 03] ont comparé les performances de l'estimateur de Wilks (au premier ordre) et d'un estimateur par tirage stratifié (à deux strates). Leur méthodologie diffère légèrement de la nôtre, mais la philosophie est similaire : quelques dizaines de calculs préliminaires permettent de spécifier les strates. Dans un travail ultérieur, il serait intéressant de comparer les estimateurs du quantile à 95% avec un niveau de confiance à 95% par les méthodes de Wilks et de stratification contrôlée avec bootstrap.

L'écart-type de l'estimateur empirique peut être réduit d'un facteur 1.5 par l'estimateur par échantillonnage préférentiel, pour notre test ( $n = 200$  et  $\alpha = 0.95$ ). Une condition très importante quant à l'obtention de bons résultats est l'existence d'une seule région d'importance pour chaque variable (ce qui doit être le cas pour Cathare). L'utilisation de mélange de densités pour la densité d'importance serait une piste intéressante pour remédier à ce problème.



## CONCLUSION

Cette conclusion propose un bilan du travail réalisé, et indique quelques perspectives d'améliorations. Les quatre premiers chapitres concernent l'estimation d'espérance mathématique et plus particulièrement l'estimation d'une probabilité de défaillance (fiabilité des structures) supposée faible. Le dernier chapitre est consacré à l'estimation de quantile.

Le premier travail effectué a été de synthétiser les méthodes et travaux existant dans ce domaine : les méthodes d'approximation (FORM/SORM), qui permettent d'estimer efficacement une probabilité faible de rupture mais malheureusement dont l'erreur ne peut être contrôlée, et les méthodes de Monte Carlo gourmandes en nombre de simulations. Pour rendre utilisable l'approche Monte Carlo, des techniques de réduction de variance sont nécessaires. Une d'elle a particulièrement retenue notre attention et a été l'objet du deuxième chapitre : le tirage d'importance, où les variables d'entrées sont tirées suivant une nouvelle densité de probabilité convenablement choisie. Nous avons aussi retenu la méthode de Monte Carlo conditionnelle (réduction de la dimension) qui peut être combinée facilement avec le tirage d'importance et qui permet un gain notable en terme de réduction de variance lorsque la variable isolée influe sur la fiabilité.

Dans le deuxième chapitre, différentes variantes du tirage d'importance dans un contexte de fiabilité des structures ont été données. Beaucoup de ces méthodes utilisent les résultats des méthodes d'approximation dans la construction de la densité d'importance ou dans l'estimateur final de la probabilité de défaillance. Nous nous sommes ensuite intéressés au tirage d'importance adaptatif : comment construire une suite de densités qui convergerait vers la densité optimale ? La méthodologie choisie consiste à se placer dans le cas paramétrique, c'est-à-dire que nous choisissons une famille de lois et c'est une suite des paramètres de cette famille qui est alors construite. Dans le cadre de l'estimation d'une probabilité de défaillance, le choix de la famille n'est pas le plus critique (sauf que son support doit inclure le support de la densité initiale) car ce sont les tirages dans les zones importantes qui nous intéressent (le domaine de défaillance en particulier) et non la forme de la densité. Pour construire cette suite, nous avons tout d'abord envisagé de choisir dans la famille une densité d'importance qui a les deux mêmes premiers moments que ceux de la densité optimale (cela peut être généralisé aux  $p$  premiers moments). Dans un deuxième temps, nous avons choisi de rechercher une suite des paramètres qui minimise la variance de l'estimateur. Pour finir, nous avons évoqué la construction d'une

suite qui minimise l'entropie croisée entre la densité d'importance choisie et la densité optimale, c'est une piste intéressante qu'il faudrait explorer. En outre, pour l'estimation de l'espérance mathématique, nous avons étudié, tout d'abord, l'estimateur classique du tirage d'importance qui utilise une densité d'importance dont les paramètres sont définis pas la dernière densité de la suite ; ensuite, un estimateur qui recycle les tirages effectués lors de la construction de la suite en se basant sur leur densité respective, et enfin, un estimateur qui utilise un mélange de toutes les densités intermédiaires (ou seulement un mélange de la densité initiale et de la dernière densité). Lorsque qu'il existe deux zones importantes, ce dernier estimateur peut permettre de ne pas trop négliger l'une des zones. Lorsque le domaine de défaillance présente  $m$  zones importantes, une densité unimodale ne permet pas de tous les prendre en compte, il faudrait alors envisager de construire une densité à  $m$  modes, ce qui demande une connaissance préalable du domaine de défaillance. Par exemple, si nous connaissons les points de conception, nous pouvons construire un mélange de densités centrées aux différents points de conception. Des simulations ont indiqué que l'estimation de la variance (dans le cas d'une famille de loi normale) est souvent trop faible et nous donne des tirages trop concentrés, en l'augmentant (double de la variance de la densité initiale), cela permet de diminuer ce problème. Cependant une trop grande augmentation de la variance peut être défavorable. Nous avons aussi mis en évidence le problème des poids trop importants dans la construction de la suite : un poids trop important dans une mauvaise direction ralentit considérablement la convergence de la suite. Pour pallier cette difficulté, nous conseillons d'affecter aux poids les plus importants une plus petite valeur. Cette même technique pourrait aussi être appliquée sur l'estimation finale de l'espérance qui deviendrait alors biaisée, ceci permettrait de ne pas donner une trop grande importance à une petite fraction de l'échantillon. En dimension  $\geq 2$ , la matrice de covariance de la densité optimale peut contenir des dépendances, dans une approche paramétrique, nous pourrions construire une suite de ces dépendances en plus des deux premiers moments.

Dans le troisième chapitre, une méthode de simulation multi-niveaux utilisant les algorithmes de Monte Carlo par Chaîne de Markov est présentée. Elle peut être réalisée en complément du tirage d'importance. Pour répondre aux besoins du CEA, nous n'avons eu à disposition que quelques centaines de simulations pour estimer la probabilité d'un évènement rare, or ces méthodes sont assez gourmandes en nombres de simulations pour être efficaces, et à nombre de simulations fixé, le tirage d'importance obtient des résultats beaucoup plus satisfaisants.

Dans le quatrième chapitre, nous avons présenté en détail le combustible à particules et son comportement sous irradiation. Nous avons ensuite parcouru l'utilisation des méthodes statistiques pour l'évaluation de la probabilité de rupture de ce type de combustible dans

divers codes de calcul internationaux et nous avons décrit sommairement le code ATLAS. Une analyse de sensibilité a d'abord été réalisée pour plusieurs raisons : la première étant de mettre en évidence les paramètres dont il est nécessaire d'améliorer la connaissance et la deuxième pour déterminer les paramètres les plus influents afin de ne prendre en compte qu'un nombre limité de paramètres dans l'estimation de la probabilité de rupture. En définitive, les méthodes de tirages d'importance paramétriques adaptatifs ont été appliquées pour estimer la probabilité de rupture moyenne d'un ensemble de particules de combustible sur plusieurs cas : l'expérience HFR-EU1 et les réacteurs GT-MHR et PBMR. Les méthodes statistiques permettent d'arriver rapidement à une bonne précision. Les résultats ont été obtenus avec l'état actuel du code et devront être validés expérimentalement.

Dans le dernier chapitre, un panorama des méthodes usuelles d'estimation de quantiles a été donné. Les quantiles que nous avons envisagés ne sont pas des quantiles extrêmes (se mesurant avec des puissances de 10), mais plutôt des quantiles de l'ordre de 95%, avec des échantillons de taille de l'ordre de 200. Nous recherchons par contre une bonne précision dans l'estimation du quantile, et/ou un certain niveau de sûreté. Nous nous sommes ensuite placés dans le cas où, premièrement la variable d'intérêt  $f(X)$  est la sortie d'un gros code numérique dont le vecteur des paramètres incertains  $X$  est de grande dimension, et deuxièmement nous avons à disposition un modèle réduit  $f_r(X)$  du code. Nous avons examiné une méthode originale de stratification contrôlée. Dans ce cas, les variables d'entrée  $X_i$  sont tirées par rejet, le rejet étant basé sur les valeurs  $f_r(X_i)$ . Nous contrôlons exactement le nombre de réalisations de  $X_i$  qui tombent dans des régions bien délimitées par le modèle réduit, et nous sommes ramenés à estimer des probabilités conditionnelles. Nous avons développé une méthode de stratification contrôlée adaptative dans laquelle les réalisations des paramètres d'entrée sont échantillonnées dans des zones déterminées par le modèle réduit, et dans laquelle le nombre de réalisations allouées dans chaque zone est optimisé de manière adaptative. La réduction de variance est alors très substantielle. Enfin, nous avons examiné une méthode de tirage d'importance (cf. chapitre 2). La réduction de variance sur l'estimateur du quantile peut alors être relativement importante. Les méthodes de stratification contrôlée et du tirage d'importance ont été appliquées sur un problème fourni par le CEA (Cadarache et Grenoble) : l'estimation du quantile à 95% du premier pic de température de gaine lors d'un scénario accidentel dans le coeur d'une centrale nucléaire (APRP - GB). Ces méthodes ont donné des résultats très prometteurs.



## RÉFÉRENCES

- [AAT 89] G. L. ANG, A. H.-S. ANG & W. H. TANG. Kernel method in importance sampling density estimation. *ICOSSAR'89, the 5th International Conference on Structural Safety and Reliability*. San Francisco. pp. 1193–1200, août 1989.
- [AAT 92] G. L. ANG, A. H.-S. ANG & W. H. TANG. Optimal importance sampling density estimation. *Journal of Engineering Mechanics*. Vol. 118, n° 6, pp. 1146–1163, 1992.
- [AB 99] S. K. AU & J. L. BECK. A new adaptive importance sampling scheme for reliability calculations. *Structural Safety*. Vol. 21, pp. 135–158, 1999.
- [AB 01] S. K. AU & J. L. BECK. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*. Vol. 16, n° 4, pp. 263–277, 2001.
- [AB 03] S. K. AU & J. L. BECK. Important sampling in high dimensions. *Structural Safety*. Vol. 25, pp. 139–163, 2003.
- [ADKOS 98] R. AMBARTZUMIAN, A. DER KIUREGHIAN, V. OHANIAN & H. SUKIASIAN. Multinormal probability by sequential conditioned sampling : theory and application. *Probabilistic Engineering Mechanics*. Vol. 13, n° 4, pp. 299–308, 1998.
- [AdR 03] E. ARDILLON & E. DE ROCQUIGNY. Probabilistic estimation of safety quantiles for complex non-regular physical systems with uncertain parameters. *Proceedings of the 25th ESREDA Seminar*. Paris. 2003.
- [AW 88] A. N. AVRAMIDIS & J. R. WILSON. Correlation-induction techniques for estimating quantiles in simulation experiment. *Oper. Res.* Vol. 46, pp. 574–591, 1988.
- [Bas ] D. BASTIEN. *Réacteurs à haute température*. Techniques de l'Ingénieur, traité Génie Nucléaire, B 3 190.
- [BB 90] C. G. BUCHER & U. BOURGUND. A fast and efficient response surface approach for structural reliability problems. *Structural Safety*. Vol. 7, pp. 57–66, 1990.
- [BdC 06] P. BAZIN & A. DE CRECY. *Contribution to BEMUSE phase III*. Rapport technique, Compte rendu d'activité CEA/DEN/GRE/SSTH/LDAS DO 126 22/06/06, 2006.

- [BDM 01] N. BARTOLI & P. DEL MORAL. *Simulations et algorithmes stochastiques*. Cepaduwès, 2001.
- [Ben 91] R.G. BENNETT. Finite element stress analysis for coated particle fuel modeling under normal operating conditions. *Nuclear Technology*. Vol. 96, pp. 117–122, 1991.
- [Ber 05] M. BERVEILLER. *Eléments finis stochastiques : approches intrusive et non intrusive pour des analyses de fiabilité*. Thèse de doctorat : Université Blaise Pascal - Clermont II, 2005.
- [BF 81] P. J. BICKEL & D. A. FREEDMAN. Some asymptotic theory for the bootstrap. *Annals of Statistics*. Vol. 9, pp. 1196–1217, 1981.
- [BF 87] P. BERNARD & M. FOGLI. Une méthode de monte carlo performante pour le calcul de la probabilité de ruine. *Revue du C.T.I.C.M.* n° 4, pp. 23–29, 1987.
- [BG 97] S. P. BROOKS & A. GELMAN. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*. Vol. 7, pp. 434–455, 1997.
- [Bje 88] P. BJERAGER. Probability integration by directional simulation. *Journal of Engineering Mechanics*. Vol. 114, n° 8, pp. 1288–1302, août 1988.
- [Bje 89] P. BJERAGER. Computational mechanics of probabilistic and reliability analysis, chapitre Methods for structural reliability computations. Elmepress International, 1989.
- [BO 92] J. O. BERGER & M. S. OH. Adaptive importance sampling in Monte Carlo integration. *Journal of Statistical Computation and Simulation*. Vol. 41, pp. 143–168, 1992.
- [BO 93] J. O. BERGER & M. S. OH. Integration of multimodal functions by Monte Carlo importance sampling. *Journal of the American Statistical Association*. Vol. 88, n° 422, pp. 450–456, 1993.
- [Bon 80] K. BONGARTZ. *Status of the fuel stress and failure rate calculations at KFA*. Rapport technique 1686, KFA, 1980.
- [Bre 84] K. BREITUNG. Asymptotic approximations for multinormal integrals. *Journal of the Engineering Mechanics, American Society of Civil Engineers*. pp. 357–366, 1984.
- [Buc 88] C. G. BUCHER. Adaptive sampling - an iterative fast Monte Carlo procedure. *Structural Safety*. Vol. 5, pp. 119–126, 1988.

- [Bur 98] C. BURGES. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*. Vol. 2, 1998.
- [CDP 05] C. CRÉMONA, N. DEVICTOR & A. PATRON. *Évaluation et prédiction de la fiabilité des structures*. Cours IPSI, 2005.
- [CE 94] G. Q. CAI & I. ELIHAKOFF. Refined second-order reliability analysis. *Structural Safety*. Vol. 14, pp. 267–276, 1994.
- [CGP 03] O. CAPPÉ, A. GUILLIN & ROBERT C. P. Population Monte Carlo. A paraître dans *journal of Computational and Graphical Statistics*, 2003.
- [Che 05] Y. CHEN. Another look of rejection sampling through importance sampling. *Statistics and Probability Letters*. Vol. 72, pp. 277–283, 2005.
- [CLRH 01] P. CHAPELOT, A. LANGUILLE, X. RAEPSAET & D. HITNER. Requirements for high temperature reactor fuel particle design assessment. *Progress in Nuclear Energy*. Vol. 38, pp. 415–418, 2001.
- [Cor 69] C. A. CORNELL. A probability based structured code. *Journal of the American Concrete Institute*. Vol. 66, n° 12, pp. 974–985, 1969.
- [Dav 81] H. A. DAVID. *Order statistics*. Wiley, New York, 1981.
- [DdFG 01] A DOUCET, N. DE FREITAS & N. GORDON. *Sequentiel Monte Carlo in Pratique*. Springer, 2001.
- [Dev 96] N. DEVICTOR. *Fiabilité et mécanique : méthodes FORM/SORM et couplages avec des codes d'éléments finis par des surfaces de réponses adaptatives*. Thèse de doctorat : Université Blaise Pascal - Clermont II, 1996.
- [DFS 97] J.-J. DROESBEKE, J FINE & G. SAPORTA. *Plans d'expériences - Applications à l'entreprise*. Editions Technip, 1997.
- [Dit 79] O. DITLEVSEN. Generalized second moment reliability index. *Journal of Structural Mechanics*. Vol. 7, pp. 435–451, 1979.
- [DKD 98] A. DER KIUREGHIAN & T. DAKESSIAN. Multiple design points in first and second-order reliability. *Structural Safety*. Vol. 20, pp. 37–49, 1998.
- [DM 96] O. DITLEVSEN & H. O. MADSEN. *Structural Reliability Methods*. Wiley & Sons, 1996.
- [DM 00] N. DEVICTOR & J.M. MARTINEZ. Nonlinear regression methods in uncertainty and sensitivities studies and reliability computations. *Foresight and Precaution*. Proceedings ESREL, pp. 1633–1639, 2000.

- [DMG 90] O. DITLEVSEN, R. E. MELCHERS & H. GLUVER. General multi-dimensional probability integration by directional simulation. *Computers and Structures*. Vol. 36, pp. 355–368, 1990.
- [Duf 96] M. DUFLO. *Algorithmes stochastiques*. Springer, 1996.
- [Efr 79] B. EFRON. Bootstraps methods : another look at the jackknife. *Annals of Statistics*. Vol. 7, pp. 1–26, 1979.
- [EFS 94] I. ENEVOLDSEN, M. H. FABER & J. D. SORENSSEN. Adaptative response techniques in reliability estimation. *Structural Safety & reliability*. Edited by SHINOZUKA & YAO SCHUËLLER. pp. 1257–1264, 1994.
- [ER 93] S. ENGELUND & S. RACKWITZ. A benchmark study on importance sampling techniques in structural reliability. *Structural Safety*. Vol. 12, pp. 255–276, 1993.
- [ETL 05] C. D. EAMON, M. THOMPSON & Z. LIU. Evaluation of accuracy and efficiency of some simulation and sampling methods in structural reliability analysis. *Structural Safety*. Vol. 27, pp. 356–392, 2005.
- [Fab 01] *Fabrication du combustible HTR*, 2001. Compte-rendu de réunion du 03/05/2001 à Cadarache CR LCU01-010.
- [Far 89] L. FARAVELLI. A response surface approach for reliability analysis. *Journal of Engineering Mechanics*. Vol. 115, n° 12, pp. 2763–2781, 1989.
- [Fel 71] W. FELLER. *An introduction to probability theory and its applications*. John Wiley & Son, New York, 1971.
- [FLM<sup>+</sup> 05] M. FÜTTERER, H. LOHNER, A. MARNIER, S. DE GROOT & C. SCIOLLA. Preliminary results from an avr fuel pebble irradiation at increased temperature and burn-up in the hfr petten. *International congress on advances in nuclear power plants*. 2005. paper 5296.
- [FLS 06] K. T. FANG, R. LI & A. SUDJIANTO. *Design and modeling for computer experiments*. Chapman and Hall, 2006.
- [FR 88] M. FUJITA & R. RACKWITZ. Updating first and second order reliability estimates by importance sampling. *Structural Engineering*. Vol. 5, pp. 53–59, 1988.
- [Fre 81] D. A. FREEDMAN. Bootstrapping regression models. *Annals of Statistics*. Vol. 9, pp. 1218–1228, 1981.
- [Fu 94] G. FU. Variance reduction by truncated multimodal importance sampling. *Structural Safety*. Vol. 13, pp. 267–283, 1994.

- [Ga 02] I. GOLUBEV & AL. Mathematical model and computer code for coated particles performance at normal operating conditions. *HTR-2002*. Petten. AIEA, April 2002.
- [Gar 00] M. GARVELS. *The splitting method for rare event simulation*. Thèse de doctorat : University of Twente, The Netherlands, 2000.
- [Gew 89] J. GEWEKE. Bayesian inference in econometric model using monte carlo integration. *Econometrica*. Vol. 57, pp. 1317–1340, 1989.
- [Gew 92] J. GEWEKE. *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*. *Bayesian Statistics 4*, pp. 169–193. Bernardo J.M., Berger J.O., Dawid A.P. & Smith A.F.M., Oxford University Press, 1992.
- [GG 99] ANNE GILLE-GENEST. *Utilisation des méthodes numériques probabilistes dans les applications au domaine de la Fiabilité des Structures*. Thèse de doctorat : Université Paris VI, 1999.
- [Gly 96] P. GLYNN. Importance sampling for monte carlo estimation of quantiles. *Mathematical Methods in Stochastic Simulation and Experimental Design : Proceedings of the 2nd St. Petersburg Workshop on Simulation*. pp. 180–185, 1996.
- [GR 88] S. GOLLWITZER & R. RACKWITZ. An efficient numerical solution to the multinormal integral. *Probabilistic Engineering Mechanics*. Vol. 3, n° 2, pp. 98–101, 1988.
- [GR 92] A. GELMAN & D. B. RUBIN. Inference from iterative simulation using multiple sequences. *Statistical Science*. Vol. 7, pp. 457–511, 1992.
- [GR 96] G. H. GIVENS & A. E. RAFTERY. Local adaptive importance sampling for multivariate densities with strong nonlinear relationships. *Journal of the American Statistical Association*. Vol. 91, pp. 132–141, 1996.
- [Har 83] A. HARBITZ. Efficient and accurate probability of failure calculation by use of the importance sampling technique. *Fourth International conference on applications of statistics and probability in soil and structural engineering*. Italy. 1983.
- [Har 86] A. HARBITZ. An efficient sampling method for probability of failure calculation. *Structural Safety*. Vol. 3, pp. 109–116, 1986.
- [Hel 83] C. W. HELSTROM. Comment : distribution of quadratic forms in normal random integration - Evaluation by numerical ntegration. *Journal of Scientific Statistical Computation*. Vol. 4, n° 2, 1983.

- [HeN 98] T. C. HESTERBERG & B. L. ET NELSON. Control variates for probability and quantile estimation. *Management Science*. Vol. 44, pp. 1295–1312, 1998.
- [Hes 88] T. C. HESTERBERG. *Advances in importance sampling*, 1988. new version in 2003.
- [Hes 95] T. HESTERBERG. Weighted average importance sampling and defensive mixture distributions. *Technometrics*. Vol. 37, n° 2, pp. 185–194, 1995.
- [HH 99] J. HUH & A. HALDAR. Reliability analysis under dynamic loading. *Applications of Statistics and Probability*. Proceedings ICASP, pp. 789–796, 1999.
- [HJ 90] T. J. HASTIE & TIBSHIRANI. R. J. *Generalized additive models*. Chapman and Hall, 1990.
- [HL 74] A. M. HASOFER & N. C. LIND. Exact and invariant second moment code format. *Journal of Engineering Mechanics*. Vol. 100, pp. 111–121, 1974.
- [HN 87] J. C. HSU & B. L. NELSON. Control variates for quantile estimation. *Proceedings of the 1987 Winter Simulation Conference*. Edited by A. THESEN, H. GRANT & W. DAVID KELTON. pp. 434–444, 1987.
- [Hon 99] H.P. HONG. Simple approximations for important second-order reliability estimates. *Journal of Engineering Mechanics*. Vol. 125, n° 5, pp. trouver pages, 1999.
- [HR 81] M. HOHENBICHLER & R. RACKWITZ. Non-normal dependent vectors in structural safety. *Journal of Engineering Mechanics*. Vol. 107, n° 6, pp. 1227–1238, 1981.
- [HR 83] M. HOHENBICHLER & R. RACKWITZ. First-order concepts in system reliability. *Structural Safety*. Vol. 1, pp. 177–188, 1983.
- [HR 88] M. HOHENBICHLER & R. RACKWITZ. Improvement of second order reliability estimates by importance sampling. *Journal of Engineering Mechanics*. Vol. 114, n° 12, pp. 2195–2199, 1988.
- [HS 96] T HOMMA & A. SALTELLI. Importance measures in global sensitivity analysis of non linear models. *Reliability Engineering and System Safety*. Vol. 52, n° 1-17, 1996.
- [HTF 02] T. HASTIE, R. TIBSHIRANI & J. FRIEDMAN. *The elements of statistical learning*. Springer, 2002.

- [Hun 04] J. D. HUNN. *Results from ONRL characterization of german reference fuel from EUO 2358-2365 composite*. Rapport technique, OAK Ridge National Laboratory, 2004.
- [Hur 04] J. E. HURTADO. An examination of methods for approximating implicit limit state functions from the viewpoint of statistical learning theory. *Structural Safety*. Vol. 26, pp. 271–293, 2004.
- [Hur 07] J. E. HURTADO. Filtered importance sampling with support vector margins : a powerful technique for structural reliability analysis. *Structural Safety*. Vol. 29, pp. 2–15, 2007.
- [Ibr 91] Y. IBRAHIM. Observations of applications of importance sampling in structural reliability analysis. *Structural Safety*. Vol. 9, pp. 269–281, 1991.
- [IRG 98] S. S. ISUKAPALLI, A. ROY & P. G. GEORGOPOULOS. *Stochastic response surface methods for uncertainty propagation : application to environmental and biological systems*. Risk analysis : an international journal, 1998.
- [Jac 06] J. JACQUES. *Contributions à l'analyse de sensibilité et à l'analyse discriminante généralisée*. Thèse de doctorat : Université Joseph Fourier - Grenoble 1, 2006.
- [Joe 95] H. JOE. Approximations to multivariate normal rectangle probabilities based on conditional expectations. *Journal of American Statistical Association*. Vol. 90, n° 431, pp. 957–964, 1995.
- [Kay 05] I. KAYMAZ. Application of kriging method to structural reliability problems. *Structural Safety*. Vol. 27, pp. 133–151, 2005.
- [KBC 88] A. KARAMCHANDANI, P. BJERAGER & C. A. CORNELL. Adaptive importance sampling. *Reliability and risk analysis in civil engineering*. 5<sup>th</sup> international conference on structural safety and reliability, pp. 930–936, 1988.
- [KN 97] S. KIM & S. NA. Response surface method using vector projected sampling points. *Structural Safety*. Vol. 19, n° 1, pp. 3–19, 1997.
- [Lag 06] A. LAGNOUX. *Analyse des modèles avec duplication des trajectoires pour l'étude des évènements rares*. Thèse de doctorat : Université Paul Sabatier Toulouse III, 2006.
- [Lai 05] M. LAINET. *Presentation of ATLAS V2.0*. Rapport technique, CEA/DEN/CAD/DEC/SESC/LSC, 2005.
- [Lan 02] A. LANGUILLE. *HTR coated particles and fuel elements - present development*, 2002. High temperature reactor school.

- [LDK 86] P. L. LIU & A. DER KIUREGHIAN. Structural reliability under incomplete probability information. *Journal of Engineering Mechanics, American Society of Civil Engineers*. Vol. 112, n° 1, pp. 85–104, 1986.
- [LDKH 87] P. L. LIU, A. DER KIUREGHIAN & S. J. HWANG. Second-order reliability approximations. *Journal of Engineering Mechanics, American Society of Civil Engineers*. Vol. 113, n° 8, pp. 1208–1225, 1987.
- [Lem 05] M. LEMAIRE. *Fiabilité des structures - Couplage mécano-fiabiliste statique*. Lavoisier Hermès, 2005.
- [Lin 93] D. LIN. A new class of super saturated designs. *Technometrics*. Vol. 35, pp. 28–31, 1993.
- [Liu 01] J. S. LIU. *Monte Carlo strategies in Scientific Computing*. Series in Statistics. Springer Verlag, 2001.
- [LP 02] A. LANGUILLE & M. PELLETIER. Comportement du combustibles RHT en réacteur. *Revue Générale Nucléaire*. n° 6, pp. 38–42, 2002.
- [LSE 02] M. LECOMTE, J. P. SERPANTIE & S. EHSTER. Les réacteurs à haute température de nouvelle génération : les options et leurs motivations. *Revue Générale Nucléaire*. n° 6, pp. 31–37, 2002.
- [LSSQ 02] M. LABAR, W. SIMON, A. SHENOY & T. QUINN. Design status of the gas turbine - modular helium reactor : An advanced nuclear power system for near term deployment. *Americas Nuclear Energy Symposium*. 2002.
- [LWDK 91] A. M. LAW & W. D. W. D. KELTON. *Simulation modeling and analysis*. 2nd édition. McGraw-Hill, 1991.
- [Mar 73] D. G. MARTIN. An analytical method of calculating, to a reasonable accuracy, stresses in the coatings of htr fuel particles. *Journal of Nuclear Materials*. Vol. 48, n° 1, pp. 35–46, 1973.
- [Mar 79] D. MARTIN. Some calculations of the failure statistics of coated fuel particles. *Nuclear Technology*. Vol. 42, pp. 304–311, 1979.
- [Mar 02] D. MARTIN. Considerations pertaining to the achievement of high burn-ups in htr fuel. *Nuclear Engineering and Design*. Vol. 213, pp. 241–258, 2002.
- [MB 93] G.K. MILLER & R.G. BENNETT. Analytical solution for stresses in triso-coated particles. *Journal of Nuclear Materials*. Vol. 206, pp. 35–49, 1993.

- [MBC 79] M. D. MACKEY, R. J. BECKAM & W. J. CONOVER. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*. Vol. 2, pp. 239–245, 1979.
- [MBD 93] M. A. MAES, K. BREITUNG & D. J. DUPUIS. Asymptotic importance sampling. *Structural Safety*. Vol. 12, pp. 167–186, 1993.
- [MBL 95] J.C. MITTEAU, A. BEAKOU & M. LEMAIRE. Less approximations for some integrals. *Applications of statistic and probability*. Edited by M LEMAIRE, J.L. FAVRE & A. MÉBARKI. Vol. 1, pp. 1003–1010, 1995.
- [Mel 89] R. E. MELCHERS. Importance sampling in structural systems. *Structural Safety*. Vol. 6, pp. 3–10, 1989.
- [Mel 90] R. E. MELCHERS. Radial importance sampling for structural reliability. *Journal of Engineering Mechanics*. Vol. 116, n° 1, pp. 189–203, 1990.
- [Mel 94] R. E. MELCHERS. Structural system reliability assessment using directional simulation. *Structural Safety*. Vol. 16, n° 1-2, pp. 23–37, 1994.
- [Mel 99] R. E. MELCHERS. *Structural reliability analysis and prediction*. 2<sup>de</sup> édition. John Wiley & Sons, 1999.
- [Mic 03] F. MICHEL. *Notice d'utilisation d'ATLAS V1.0*. Rapport technique, CEA/DEN/CAD/DEC/SESC/LSC, 2003.
- [Mit 99] J.C. MITTEAU. Error evaluation for the computation of failure probability in static structural reliability problems. *Probabilistic Engineering Mechanics*. Vol. 14, pp. 119–135, 1999.
- [MK 03] Y. MORI & T. KATO. Multinormal integrals by importance sampling for series system reliability. *Structural Safety*. Vol. 25, pp. 363–378, 2003.
- [MKL 86] H. O. MADSEN, S KRENK & N. C. LIND. *Methods of Structural Safety*. Prentice-Hall Inc., Englewood Cliffs, 1986.
- [ML 99] A. MOHAMED & M. LEMAIRE. Improved response surface method by using order sensitivity operators. *Application of statistics and probability*. Proceedings ICASP, pp. 117–124, 1999.
- [MLBL 93] J. P. MUZEAU, M LEMAIRE, P. BESSE & J. M. LOCC. Evaluation of reliability in case of complex mechanical behavior. *12<sup>th</sup> international Conference OMAE'93*. Vol. 2, 1993.

- [MLET 92] J. P. MUZEAU, M. LEMAIRE & K. EL-TAWIL. Méthodes fiabilistes des surfaces de réponses quadratiques (srq) et évaluation des règlements. *Construction Métalliques*. n° 3, pp. 41–52, 1992.
- [Mor 91] M. D. MORRIS. Factorial sampling plans for preliminary computational experiments. *Technometrics*. Vol. 33, pp. 161–174, 1991.
- [MPM 02] G.K. MILLER, D.A. PETTI & J.T. MAKI. development of an integrated performance model for triso-coated gas reactor particle fuel. *HTR-2002*. Petten. AIEA, April 2002.
- [MPVM 03] G. K. MILLER, D. A. PETTI, D. J. VARACALLE & J. T. MAKI. Statistical approach and benchmarking for modeling of multi-dimensional behavior in TRISO-coated fuel particles. *Journal of Nuclear Materials*. Vol. 317, pp. 69–82, 2003.
- [MW 94] G.K. MILLER & D.C. WADSWORTH. Treating asphericity in fuel particle pressure vessel modeling. *Journal of Nuclear Materials*. Vol. 211, pp. 57–69, 1994.
- [Nat 62] A. NATAF. Détermination des distributions dont les marges sont données. *Comptes rendus de l'Académie des Sciences de Paris*. Vol. 255, pp. 42–43, 1962.
- [NE 00] J. NIE & B. R. ELLINGWOOD. Directional methods for structural reliability analysis. *Structural Safety*. Vol. 22, pp. 233–249, 2000.
- [Nel 80] B. L. NELSON. Control variates remedies. *Operations Research*. Vol. 38, pp. 974–992, 1980.
- [NeWGB 04] W. T. NUTT & ET WALLIS G. B. Evaluation of nuclear safety from the outputs of computer codes in the presence of uncertainties. *Reliability Engineering and System Safety*. Vol. 83, pp. 57–77, 2004.
- [Nie 92] H. NIEDERREITER. Random number generation and quasi Monte Carlo methods. *Society for Industrial and Applied Mathematics*. 1992.
- [Oak 04] J. OAKLEY. Estimating percentiles of uncertain computer code outputs. *Appl. Statist.* Vol. 53, pp. 83–93, 2004.
- [OSD 03] A. OLSSON, G. SANDBERG & O DAHLBLOM. On latin hypercube sampling for structural reliability analysis. *Structural Safety*. Vol. 25, pp. 47–68, 2003.
- [PBP 99] D. C. POLIDORI, J. L. BECK & C. PAPADIMITRIOU. Nex approximations for reliability integrals. *Journal of Engineering Mechanics*. Vol. 125, n° 4, pp. 466–475, 1999.

- [PC 05] G. PUJOL & C. CANNAMÉLA. *Analyse de sensibilité du code de calcul ATLAS*. Rapport technique Note technique SESC/LSC 05-034, CEA, 2005.
- [Pen 00] M. PENDOLA. *Fiabilité des structures en contexte d'incertitudes statistiques et d'écarts à la modélisation*. Thèse de doctorat : Université Blaise Pascal - Clermont II, 2000.
- [Phe 05] M. PHELIP. European programme on high and very high temperature reactor fuel technology. *International congress on advances in nuclear power plants*. 2005. paper 5160.
- [Phi 97] A. PHILIPPE. Simulation output by riemann sums. *Journal of Statistical Computational Simulation*. Vol. 59, n° 4, pp. 295–314, 1997.
- [PLT 98] M. PAPADRAKAKIS, N. D. LAGAROS & Y. TSOMPANAKIS. Structural optimization using evolution strategies and neural networks. *Comput. Methods Appl. Mech. Engineering*. Vol. 124, n° 2, pp. mettre pages, 1998.
- [PS 02] M. D. PANDLEY & A. SARKAR. Comparison of a simple approximation for multinormal integration with an importance sampling based simulation method. *Probabilistic Engineering Mechanics*. Vol. 17, pp. 215–218, 2002.
- [Rac 01] R. RACKWITZ. Reliability anlysis - a review and some perspectives. *Structural Safety*. Vol. 23, pp. 365–395, 2001.
- [Rao 73] C. R. RAO. *Linear statistical inference and its applications*. Wiley, 1973.
- [RC 99] C. ROBERT & G. CASELLA. *Monte Carlo statistical methods*. Springer-Verlag, New York, 1999.
- [RE 93] M.R. RAJASHEKHAR & B. R. ELLINGWOOD. A response surface approach for reliability analysis. *Structural Safety*. Vol. 12, pp. 205–220, 1993.
- [Rei 04] F. REITSMA. The pebble bed modular reactor layout and neutronics design of the equilibrium cycle. *PHYSOR- the Physics of Fuel Cycles and Advanced Nuclear Systems : Global Developments*. Chicago. 2004.
- [RF 78] R. RACKWITZ & B. FIESSLER. Structural reliability under combined load sequences. *Computer & Structures*. Vol. 9, pp. 489–494, 1978.
- [Ric 80] S. O. RICE. Distribution of quadratic forms in normal random variables - Evaluation by numerical integration. *Journal of Scientific Statistical Computation*. Vol. 1, n° 4, 1980.

- [RK 04] R. Y. RUBINSTEIN & D. P. KROESE. *The cross entropy method - A unified approach for combinatorial optimization, Monte Carlo simulation and machine learning*. Springer Verlag, 2004.
- [RL 92] A. E. RAFTERY & S. LEWIS. How many iterations in the gibbs sampler? *Bayesian Statistics 4*. Oxford. pp. 763–773, 1992.
- [Ros 52] M. ROSENBLATT. Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*. Vol. 23, pp. 470–472, 1952.
- [Rub 81] R. Y. RUBINSTEIN. *Simulation and the Monte Carlo Methods*. Probability and Statistics. John Wiley & Sons, 1981.
- [Rut 06] B. RUTHERFORD. A response-modeling alternative to surrogate models for support in computational analyses. *Reliability Engineering and System Safety*. Vol. 91, pp. 1322–1330, 2006.
- [Sa 01] K. SAWA & AL. Prediction of fuel performance and fission gas release behavior during normal operation of the high temperature engineering test reactor by JAERI and FZJ modeling approach. *Nuclear Science and Technology*. Vol. 38, n° 6, pp. 411–419, 2001.
- [Sat 00] F. SATTERTHWAITTE. Random balance experimentation. *Technometrics*. pp. 111–137, 0000.
- [SCS 00] A. SALTELLI, K. CHAN & E.M. SCOTT. *Sensitivity Analysis*. Series in Probability and Statistics. Wiley, 2000.
- [Shi 83] M. SHINOZUKA. Basic analysis of structural safety. *Journal of Structural Engineering*. Vol. 109, pp. 721–740, 1983.
- [SM 93] A. SELLIER & A. MEBARKI. Evaluation de la probabilité d'occurrence d'un évènement rare par utilisation du tirage d'importance conditionné. *Annales des Ponts et Chaussées*. Vol. 3<sup>me</sup> trimestre, 1993.
- [SMW 89] W. J. SACKS, J. AND. WELCH, T. J. MITCHELL & H. P. WYNN. Design and analysis of computer experiments. *Statistical Science*. Vol. 4, n° 409-435, 1989.
- [SS 87] G. I. SCHULLER & R. STIX. A critical appraisal of methods to determine failure probabilities. *Structural Safety*. Vol. 4, pp. 293–309, 1987.
- [SS 96] K. SAWA & S. SHIOZAWA. development of a coated fuel particle failure model under high burnup irradiation. *Nuclear Science and Technology*. Vol. 33, n° 9, pp. 712–720, 1996.

- [STC 99] A. SALTELLI, S. TARANTOLA & K. CHAN. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*. Vol. 41, pp. 39–56, 1999.
- [STCR 04] A. SALTELLI, S. TARANTOLA, F. CAMPOLONGO & M. RATTO. *Sensitivity Analysis in Practice - a guide to Assessing Scientific Models*. Wiley, 2004.
- [Ste 87] M. STEIN. Large sample properties of simulations using latin hypercube sampling. *Technometrics*. Vol. 29, n° 2, pp. 143–151, 1987.
- [TH 00] A. TAKEMURA & N. HOSHINO. On reduction of finite-sample variance by extended latin hypercube sampling. *Bernoulli*. Vol. 6, n° 6, pp. 1035–1050, 2000. téléchargeable sur la page personnelle d’Akimichi Takemura.
- [Tuf 97] B. TUFFIN. *Simulation accélérée par les méthodes de Monte Carlo et quasi Monte Carlo : théories et applications*. Thèse de doctorat : Université de Rennes 1, 1997.
- [Tve 90] L. TVEDT. Distribution of quadratic forms in normal space - Application to structural reliability. *Journal of Engineering Mechanics*. Vol. 116, n° 6, pp. 1183–1197, 1990.
- [VAVA 91] M. VILLÉN-ALTAMIRANO & J. VILLÉN-ALTAMIRANO. Restart : a method for accelerating rare event simulation. *Queueing, Performance and Control in ATM : Proceedings of the 13rd International Teletraffic Congress*. Edited by J. W. COHEN & C. D. PACK. Numéro 15, Amsterdam. pp. 71–76, 1991.
- [VAVA 94] M. VILLÉN-ALTAMIRANO & J. VILLÉN-ALTAMIRANO. Restart : a straightforward method for fast simulation of rare events. *Proceedings of the 1994 Winter Simulation Conference*. Edited by J. D. TEW, M. S. MANIVANNAN, D. A. SADOWSKI & A. F. SEILA. Orlando. pp. 282–289, 1994.
- [VN 90] K. VERFONDERN & H. NABIELEK. *The mathematical basis of the PANAMA-I code for modeling pressure vessel failure of TRISO coated particles under accident conditions*. Rapport technique, KFA, 1990.
- [VPM ] E. VAZQUEZ & M. PIERA MARTINEZ. Estimation of the volume of an excursion set of a gaussian process using intrinsic kriging. soumis à *Computational Statistics and Data Analysis*.
- [Wal 72] H. WALTHER. On mathematical models for calculating the mechanical behaviour of coated fuel particles. *Nuclear Engineering and design*. Vol. 18, pp. 11–39, 1972.

- [WAL 94] G. L. WANG, AHS ANG & J. C. LEE. Adaptive kernel method for evaluating structural system reliability. *Structural Engineering and Mechanics*. Vol. 5, n° 2, 1994.
- [WB 03] J. WANG & R.G. BALLINGER. An improved coated particle fuel failure model. *Global 2003*. New Orleans. pp. 531–536, November 2003.
- [WBD 04] J. WANG, R.G. BALLINGER & J.T. DIECKER. design optimization and analysis of coated particle fuel using advanced fuel performance modelling techniques. *2nd international topical meeting on HTR technology*. Beijing. september 2004.
- [WBM 04] J. WANG, R.G. BALLINGER & H.J. MACLEAN. TIMCOAT : an integrated fuel performance model for coated particle fuel. *Materials for Nuclear Systems*. Vol. 148, pp. 68–96, 2004.
- [WH 71] G.K. WILLIAMSON & P. HORNER. A simple model describing the physical processes which determine coated particle life. *3rd conf. Industrial Carbons and Graphite*. London. pp. 383, 1971.
- [Wu 92] Y. T. WU. An adaptive importance sampling method for structural system reliability analysis. *Reliability Technology*. Vol. 28, pp. 217–231, 1992.
- [ZDK 94] Y ZHANG & A. DER KIUREGHIAN. Two improved algorithms for reliability analysis. *Reliability and optimization of structural systems*. Edited by R RACKWITZ & BORRI A. AUGUSTI, G. 6<sup>th</sup> IFIP WG 7.5 working conference on reliability of structural systems, pp. 297–304, 1994.
- [Zha 96] P. ZHANG. Nonparametric importance sampling. *Journal of the American Statistical Association*. Vol. 91, n° 435, pp. 1245–1253, Septembre 1996.
- [ZO 99a] Y. G. ZHAO & T. ONO. New approximations for SORM : Part 1. *Journal of Engineering Mechanics*. Vol. 125, n° 1, pp. 79–85, 1999.
- [ZO 99b] Y. G. ZHAO & T. ONO. New approximations for SORM : Part 2. *Journal of Engineering Mechanics*. Vol. 125, n° 1, pp. 86–93, 1999.

## ANNEXE A

## MÉTHODES D'ANALYSE DE SENSIBILITÉ

Cette partie est extraite de [PC 05].

## A.1 Étude d'un modèle linéaire

Dans cette section, nous nous intéressons à un modèle linéaire de la forme

$$Y = b_0 + \sum_{j=1}^p b_j X_j + \varepsilon$$

Dans la pratique nous ne savons pas à priori si le modèle est linéaire. L'analyse est donc faite sur une régression linéaire du modèle, si cette régression est de bonne qualité. Dans le cas où la régression linéaire n'est pas correcte, l'information des indices peut quand même être prise en compte si elle est considérée qualitativement (classement des facteurs par ordre d'importance).

La méthode consiste à estimer différents indices d'influence à partir d'un échantillon des facteurs

$$X = \begin{pmatrix} X_{11} & \dots & X_{1p} \\ \vdots & & \vdots \\ X_{n1} & \dots & X_{np} \end{pmatrix}$$

## Coefficients de corrélation (PEAR)

Le coefficient de corrélation de Pearson entre le facteur  $X_j$  et la réponse  $Y$  est donné par

$$\text{PEAR}_j = \widehat{\text{corr}}(X_j, Y) = \frac{\sum_{i=1}^n (X_{ij} - \mathbb{E}(X_j))(Y_i - \mathbb{E}(Y))}{\left[ \sum_{i=1}^n (X_{ij} - \mathbb{E}(X_j))^2 \right]^{1/2} \left[ \sum_{i=1}^n (Y_i - \mathbb{E}(Y))^2 \right]^{1/2}}$$

Il mesure la qualité de la relation linéaire entre les variables  $X_j$  et  $Y$ . Il est égal à  $\pm 1$  si et seulement si  $Y = aX_j + b$ , et si  $X_j$  et  $Y$  sont indépendantes alors il est nul (la réciproque est fautive en général). Un signe positif indique que les variables progressent dans le même sens, et un signe négatif dans le sens contraire.

### Coefficient de régression standards (SRC)

L'indice SRC (Standardized Regression Coefficient) correspondant au facteur  $X_j$  est défini par

$$\text{SRC}_j = \frac{\text{Var}(X_j)}{\text{Var}(Y)} b_j^2$$

Les indices SRC sont les carrés des coefficients du modèle de régression linéaire dans lequel les facteurs et la réponse ont été standardisés. Ils mesurent donc l'importance d'une variable d'entrée sur la variable de sortie. Si les variables d'entrée sont indépendantes, leur somme vaut 1 ; ils forment alors une décomposition de la variance de la réponse. Un indice SRC va dans ce cas mesurer la part de la variance de la réponse expliquée par un facteur.

La théorie de la régression linéaire apporte que les indices SRC sont les carrés des coefficients PEAR. Nous conservons cependant les deux classes d'indices . Les indices PEAR donnent une information sur le signe, donc sur le sens de l'influence. Quand aux indices SRC, ils quantifient l'influence de par la propriété de décomposition de la variance.

### Coefficients de corrélation partielle (PCC)

Dans le cas où les facteurs sont corrélés les indices SRC perdent une grande partie de leur pertinence. Par exemple dans le cas de trois variables corrélées deux à deux  $X_1$ ,  $X_2$  et  $Y$ , ils ne permettent pas de déterminer l'influence de  $X_1$  seul sur  $Y$  car les effets de  $X_2$  sont implicites. Les coefficients de corrélation partielle (PCC : Partial Correlation Coefficient) sont introduits pour remédier à cela. Ils sont définis par

$$\text{PCC}_j = \text{corr}(Y - \hat{Y}, X_j - \hat{X}_j)$$

où  $\hat{Y}$  est la prévision du modèle linéaire dans lequel  $X_j$  n'est pas présent, ce modèle s'écrit

$$Y = c_0 + \sum_{\substack{1 \leq k \leq p \\ k \neq j}} c_k X_k + \varepsilon_1$$

et  $\hat{X}_j$  est la prévision du modèle linéaire qui exprime  $X_j$  en fonction des autres facteurs, ce modèle s'écrit

$$X_j = d_0 + \sum_{\substack{1 \leq k \leq p \\ k \neq j}} d_k X_k + \varepsilon_2$$

Les PCC mesurent la corrélation linéaire entre une réponse et un facteur après avoir effectué une correction visant à neutraliser les corrélations (linéaires) entre les variables d'entrée. Ils sont compris entre -1 et 1 et s'interprètent comme des coefficients de corrélation. Ce n'est pas à proprement parler une mesure d'influence, mais plutôt une mesure de la linéarité. L'interprétation est délicate et il est préférable de considérer cette mesure qualitativement.

## Analyse sur les rangs

Lorsque la relation entre  $Y$  et les  $X_i$  n'est pas linéaire mais monotone, l'analyse précédente peut être réalisée sur les rangs. Au lieu de considérer les valeurs des réalisations de  $Y$  et  $X_i$ , nous calculons leurs rangs : 1 pour la valeur la plus petite,  $\dots$ ,  $n$  pour la plus grande. En effet une relation monotone par rapport à chaque facteur se traduit par une relation linéaire sur les rangs. Les mêmes indices sont alors calculés, seuls les noms changent :

- les coefficients de corrélation de PEARSON deviennent ceux de SPEARMAN ;
- les SRC deviennent les SRRC (Standardized Rank Regression Coefficient) ;
- les PCC deviennent les PRCC (Partial Rank Correlation Coefficient).

L'interprétation est plus délicate car le modèle n'est plus étudié directement. L'information est donc à considérer qualitativement.

## A.2 Méthode de « screening » de Morris

L'expérience montre que dans le cas d'un modèle avec beaucoup de facteurs, seulement quelque-uns sont influents. Les méthodes de « screening » visent à déterminer rapidement ces facteurs. Elles reposent sur des plans d'expériences et fournissent une information qualitative, c'est-à-dire qu'elles permettent de classer les facteurs par ordre d'importance sans pour autant la quantifier. Une étude de « screening » fournit une information précieuse à moindre coût. Elle sert souvent à « élaguer » un modèle compliqué ou coûteux en temps de calcul pour pouvoir ensuite l'étudier par des méthodes plus riches.

La méthode de MORRIS permet de classer les facteurs selon trois catégories :

1. facteurs ayant des effets négligeables ;
2. facteurs ayant des effets linéaires et sans interaction ;
3. facteurs ayant des effets non linéaire et/ou avec interactions.

## Effets élémentaires

Nous supposons que tous les facteurs suivent une loi uniforme  $\mathcal{U}(0, 1)$ . L'espace des facteurs  $[0, 1]^p$  est discrétisé selon une grille  $p$ -dimensionnelle  $\{0, \frac{1}{n-1}, \frac{2}{n-1}, \dots, \frac{n-2}{n-1}, 1\}^p$ . Soit  $\Delta$  un multiple positif de  $\frac{1}{n-1}$ . Le  $i^e$  effet élémentaire est la variation du modèle perturbé dans la  $i^e$  direction, soit

$$d_i(x) = \frac{y(x + \Delta e_i) - y(x)}{\Delta}$$

où  $e_i$  est le  $i^e$  vecteur de la base canonique. Bien entendu, le point  $x$  doit être tel que  $x + \Delta e_i$  reste dans  $[0, 1]^p$ .

## Mesures de sensibilité

À partir des effets élémentaires nous définissons les mesures de sensibilité

$$\mu_i^* = \mathbb{E}(|d_i|)$$

$$\sigma_i = \sigma(d_i)$$

Ces mesures sont des mesures de sensibilité, mais au lieu d'être mesurées localement elles sont moyennées sur tout le domaine.

L'interprétation de ces quantités est assez facile :

- $\mu_i^*$  est une mesure de la sensibilité du modèle au  $i^{\text{e}}$  facteur ; en effet, si  $\mu_i^*$  est très faible, alors les effets élémentaires sont négligeables ; au contraire si  $\mu_i^*$  est grand, alors les effets élémentaires sont importants en moyenne ;
- $\sigma_i$  est une mesure de la non-linéarité ou des interactions (sans pouvoir faire de distinction entre les deux) ; en effet, si  $\sigma_i$  est petit alors les effets élémentaires sont significativement identiques : ils sont linéaires et sans interaction (les effets étant similaires, ils ne dépendent pas de la position dans l'espace, donc de la valeur des autres facteurs) ; au contraire, si  $\sigma_i$  est grand alors les effets élémentaires sont significativement différents les uns des autres : ils sont non linéaires, ou avec des interactions, ou les deux.

Un graphe  $(\mu^*, \sigma)$  permet très facilement de distinguer les différents groupes.

## Plan d'expériences

Quand le nombre de facteurs est important, les effets élémentaires ne peuvent être calculés qu'en un nombre réduit de points car le nombre d'appels au modèle doit rester raisonnable. L'ensemble des combinaisons de facteurs nécessaires pour une analyse forme donc un plan d'expériences factoriel. Le plan d'expériences de MORRIS a deux caractéristiques remarquables :

- deux simulations successives ne diffèrent que d'un facteur ; cette propriété est dénommée OAT (One At a Time) ;
- le plan d'expériences est aléatoire.

Les points de la matrice du plan d'expériences forment une *marche aléatoire* dans  $[0, 1]^p$ . C'est une matrice  $(p + 1) \times p$  telle que :

- deux lignes consécutives sont identiques sauf pour une valeur où elles diffèrent de  $\pm\Delta$  (propriété OAT) ;
- ceci se produit exactement une fois pour chaque coordonnée,

ce qui s'écrit formellement

$$B^* = (x_j^i)_{\substack{i=1\dots p+1 \\ j=1\dots p}}$$

avec

$$\begin{cases} \text{pour tout } i, \text{ il existe } j_i \text{ tel que } x^{i+1} = x^i \pm \Delta e_{j_i} \\ \text{si } j_i = j_{i'}, \text{ alors } i = i' \end{cases}$$

Les effets élémentaires sont alors donnés par

$$d_{j_i}(x^i) = \pm \frac{y(x^{i+1}) - y(x^i)}{\Delta}$$

La matrice du plan d'expérience  $B^*$  peut se construire par la formule

$$B^* = \left( J_{p+1,1} \cdot x^* + \frac{\Delta}{2} \left( (2B - J_{p+1,p})D^* + J_{p+1,p} \right) \right) P^* \quad (\text{A.2.1})$$

avec :

- $J_{i,j}$  une matrice  $i \times j$  remplie de 1 ;
- $B$  une matrice  $(p+1) \times p$  avec des 1 dans la partie triangulaire inférieure et des 0 dans la partie supérieure ;
- $D^*$  une matrice  $p \times p$  diagonale composée de +1 et -1 équiprobables ;
- $P^*$  une matrice  $p \times p$  de permutation aléatoire ;
- $x^*$  un point de la grille choisi aléatoirement ;

sous la condition que les points du plan d'expérience (les lignes de la matrice) restent dans la grille. Une méthode pour s'en assurer est de réitérer le tirage du point  $x^*$  jusqu'à vérification de la contrainte.

Les lignes  $i$  et  $i+1$  de la matrice  $B^*$  servent à calculer le  $j^{\text{e}}$  effet élémentaire, avec  $j$  tel que  $P^*(i, j) = 1$ . Le signe de cet effet (a-t-on ajouté ou retranché  $\Delta$  ?) est  $D^*(i, i)$ . En notant  $y = f(B^*)$ , le vecteur des effets élémentaires  $d$  (un effet par facteur) est donc donné par :

$$d = \left( \frac{(y_2 \cdots y_{p+1}) - (y_1 \cdots y_p)}{\Delta} \right) \cdot D^* \cdot P^*$$

En construisant  $r$  plans d'expérience nous obtenons  $r$  vecteurs  $d$ . Ces vecteurs nous donnent  $r$  effets élémentaires pour chaque facteur, c'est à dire un  $r$ -échantillon de la loi de  $d_i$  pour chaque  $i$ . Les statistiques  $\mu_i^*$  et  $\sigma_i$  s'estiment facilement sur cet échantillon par les estimateurs classiques de la moyenne et de l'écart-type.

### Choix des paramètres

Un point important est le choix des paramètres  $n$ ,  $\Delta$  et  $r$ . SALTELLI (en suivant MORRIS) conseille  $\Delta = n/(2(n-1))$  avec  $n$  pair. Nous préférons une valeur faible de  $\Delta$  (de l'ordre de  $1/(n-1)$ ) avec  $n$  assez grand. Les effets élémentaires représentent alors des variations plus locales. Mais le nombre de répétitions  $r$  requises pour explorer suffisamment l'espace est plus important. Le choix de  $\Delta$  et de  $n$  dépend donc du nombre de simulations que l'on veut faire (ce nombre est égal à  $(p+1) \times r$ ).

### A.3 Méthode de Sobol

La méthode de SOBOL est la méthode la plus générale d'analyse d'influence reposant sur une décomposition de la variance. Quatre familles d'indices sont mis à notre disposition :

- les indices d'ordre 1 expriment l'influence d'un facteur seul ;
- les indices d'ordres supérieurs expriment l'influence de l'interaction de plusieurs facteurs ;
- les indices totaux expriment l'influence totale d'un facteur, interactions comprises ;
- les indices multidimensionnels permettent les analyses précédentes dans le cas où les facteurs sont corrélés.

#### Indices de Sobol

Pour une fonction  $f$  intégrable sur  $[0, 1]^p$  nous avons la décomposition unique

$$Y = f(X_1, \dots, X_p) = f_0 + \sum_{1 \leq i \leq p} f_i(X_i) + \sum_{1 \leq i < j \leq p} f_{ij}(X_i, X_j) + \dots + f_{12\dots p}(X_1, \dots, X_p)$$

où  $f_0$  est une constante et

$$\int_0^1 f_{i_1 \dots i_s}(x_{i_1}, \dots, x_{i_s}) dx_{i_k} = 0, \quad \forall \{i_1, \dots, i_s\} \subseteq \{1, \dots, p\}, \quad \forall k = 1, \dots, s$$

La variance se décompose aussi en sommes

$$\text{Var}(Y) = \sum_{i=1}^p V_i + \sum_{1 \leq i < j \leq p} V_{ij} + \sum_{1 \leq i < j < k \leq p} V_{ijk} + \dots$$

avec

$$V_i = \text{Var}(\mathbb{E}(Y/X_i))$$

$$V_{ij} = \text{Var}(\mathbb{E}(Y/X_i, X_j)) - V_i - V_j$$

$$V_{ijk} = \text{Var}(\mathbb{E}(Y/X_i, X_j, X_k)) - V_{ij} - V_{ik} - V_{jk} - V_i - V_j - V_k$$

...

Si les facteurs  $X_i$  sont indépendants alors

$$V_{i_1 i_2 \dots i_s} = \text{Var}(f_{i_1 i_2 \dots i_s}(X_{i_1}, X_{i_2}, \dots, X_{i_s}))$$

et les indices d'influence sont donnés par

$$S_{i_1 i_2 \dots i_s} = \frac{V_{i_1 i_2 \dots i_s}}{\text{Var}(Y)}$$

Les indices d'influence traduisent la part de la variance de  $Y$  due à une variable d'entrée seule.

Les indices d'influence d'ordre 2 traduisent la part de la variance de  $Y$  due à l'interaction

de deux variables d'entrée (et non prise en compte par les variables seules). Et ainsi de suite pour les ordres supérieurs ... Le nombre de ces indices augmente de manière factorielle avec le nombre de facteurs et dans la pratique il est déjà difficile d'estimer tous les indices jusqu'à l'ordre 2.

### Indices totaux

Les indices d'influence totale traduisent la part de la variance de  $Y$  due à toutes les contributions d'un facteur. Il suffit pour cela de sommer tous les indices se rapportant à un facteur, c'est-à-dire

$$S_i^{\text{tot}} = \sum_{K \ni i} S_K$$

Dans la pratique nous verrons que nous n'aurons pas besoin des indices de tous les ordres pour les calculer. Ils s'estiment aussi simplement que les indices d'ordre 1. Ce sont donc des indices très utiles car ils fournissent une information complète sur un facteur à moindre coût.

### Indices multidimensionnels

La thèse de Julien JACQUES [Jac 06] met en évidence que dans le cas des facteurs corrélés, les indices précédents n'ont plus beaucoup de sens. Il propose de regrouper les variables corrélées entre elles et de calculer les indices entre les groupes de corrélation. Plus précisément, de nouvelles variables sont introduites (après renumérotation),

$$\begin{aligned}\tilde{X}_1 &= (X_1, \dots, X_{n1}) \\ \tilde{X}_2 &= (X_{n1+1}, \dots, X_{n2}) \\ &\dots\end{aligned}$$

avec les  $\tilde{X}_i$  indépendantes deux à deux. Nous retombons alors sur le cas des facteurs indépendants avec les indices que nous venons de présenter. Ainsi par exemple l'indice d'ordre 1 du couple de variables corrélées  $(X_i, X_j)$  est donné par

$$S_{\{i,j\}} = \frac{V_{\{i,j\}}}{\text{Var}(Y)}$$

où

$$V_{\{i,j\}} = \text{Var}(\mathbb{E}(Y/(X_i, X_j)))$$

### Estimation

Pour estimer  $V_i$  remarquons que la quantité  $\text{Var}(\mathbb{E}(Y/X_i))$  s'écrit sous la forme

$$\begin{aligned}\text{Var}(\mathbb{E}(Y/X_i)) &= \mathbb{E}(\mathbb{E}(Y/X_i)^2) - \mathbb{E}(\mathbb{E}(Y/X_i))^2 \\ &= \mathbb{E}(\mathbb{E}(Y/X_i)^2) - \mathbb{E}(Y)^2\end{aligned}$$

Le premier terme donne

$$\begin{aligned}\mathbb{E}(\mathbb{E}(Y/X_i)^2) &= \int \left( \int f(x_1, \dots, x_p) \prod_{\substack{j=1 \dots n \\ j \neq i}} dX_j \right)^2 dX_i \\ &= \int f(x_1^{(1)}, \dots, x_p^{(1)}) f(x_1^{(2)}, \dots, x_i^{(1)}, \dots, x_p^{(2)}) \prod_{i=1 \dots n} dX_i^{(1)} \prod_{\substack{j=1 \dots n \\ j \neq i}} dX_j^{(2)}\end{aligned}$$

Cette dernière intégrale s'estime par la somme

$$\mathbb{E}(\mathbb{E}(Y/X_i)^2) \simeq \frac{1}{n} \sum_{k=1}^n f(x_{k1}^{(1)}, \dots, x_{kp}^{(1)}) f(x_{k1}^{(2)}, \dots, x_{ki}^{(1)}, \dots, x_{kp}^{(2)})$$

qui nécessite deux réalisations indépendantes du  $n$ -échantillon

$$x^{(1)} = \begin{pmatrix} x_{11}^{(1)} & \dots & x_{1p}^{(1)} \\ \vdots & & \vdots \\ x_{n1}^{(1)} & \dots & x_{np}^{(1)} \end{pmatrix} \quad x^{(2)} = \begin{pmatrix} x_{11}^{(2)} & \dots & x_{1p}^{(2)} \\ \vdots & & \vdots \\ x_{n1}^{(2)} & \dots & x_{np}^{(2)} \end{pmatrix}$$

L'estimateur de  $S_i$  est donc

$$\hat{S}_i = \frac{\hat{V}_i}{\widehat{\text{Var}}(Y)}$$

où

$$\hat{V}_i = \frac{1}{n} \sum_{k=1}^n f(x_{k1}^{(1)}, \dots, x_{kp}^{(1)}) f(x_{k1}^{(2)}, \dots, x_{ki}^{(1)}, \dots, x_{kp}^{(2)}) - \widehat{\mathbb{E}(Y)}^2$$

Les autres indices s'estiment de manière similaire (sans justification) par

$$\hat{S}_{ij} = \frac{\hat{V}_{ij}}{\widehat{\text{Var}}(Y)}, \quad \dots, \quad \hat{S}_i^{\text{tot}} = 1 - \frac{\hat{V}_{\sim i}}{\widehat{\text{Var}}(Y)}, \quad \hat{S}_{\{i,j\}} = \frac{\hat{V}_{\{i,j\}}}{\widehat{\text{Var}}(Y)}, \quad \dots$$

où

$$\hat{V}_{ij} = \frac{1}{n} \sum_{k=1}^n f(x_{k1}^{(1)}, \dots, x_{kp}^{(1)}) f(x_{k1}^{(2)}, \dots, x_{ki}^{(1)}, \dots, x_{kj}^{(1)}, \dots, x_{kp}^{(2)}) - \widehat{\mathbb{E}(Y)}^2 - \hat{V}_i - \hat{V}_j$$

...

$$\hat{V}_{\sim i} = \frac{1}{n} \sum_{k=1}^n f(x_{k1}^{(1)}, \dots, x_{kp}^{(1)}) f(x_{k1}^{(2)}, \dots, x_{ki}^{(2)}, \dots, x_{kp}^{(1)}) - \widehat{\mathbb{E}(Y)}^2$$

$$\hat{V}_{\{i,j\}} = \frac{1}{n} \sum_{k=1}^n f(x_{k1}^{(1)}, \dots, x_{kp}^{(1)}) f(x_{k1}^{(2)}, \dots, x_{ki}^{(1)}, \dots, x_{kj}^{(1)}, \dots, x_{kp}^{(2)}) - \widehat{\mathbb{E}(Y)}^2$$

...

Il faut deux réalisations du  $n$ -échantillon pour estimer un indice, soit  $2n$  simulations. Les mêmes échantillons peuvent servir pour estimer les indices de tous les ordres, mais les substitutions des colonnes impliquent de nouvelles simulations. Il faut donc au total  $(p+1) \times n$  simulations pour estimer les indices d'ordre 1 ou les indices totaux. Dans la pratique  $n$  peut valoir facilement 10 000. La méthode est donc réservée aux fonctions analytiques ainsi qu'aux surfaces de réponse plutôt qu'aux codes de calcul.

#### A.4 Méthode FAST

La méthode FAST (Fourier Amplitude Sensitivity Test) repose sur une idée simple : en faisant osciller les facteurs, chacun suivant sa fréquence propre, la réponse va osciller suivant les fréquences des facteurs influents. Cette méthode a été découverte par CUKIER en 1973, et améliorée récemment par SALTELLI [STC 99].

#### Transformations

Nous supposons que les facteurs suivent des lois uniformes  $\mathcal{U}(0,1)$ . Ils sont donnés par les transformations

$$X_i(S) = g_i(\sin(\omega_i S)), \quad S \hookrightarrow \mathcal{U}(-\pi, \pi)$$

où les  $\omega_i$  sont les fréquences. Ces transformations impliquent que les facteurs ne suivront pas des lois uniformes, et qu'ils ne seront pas indépendants. Le choix des fonctions  $g_i$  et des fréquences  $\omega_i$  doit être fait pour s'en approcher au mieux. Une transformation qui garantit assez bien l'uniformité est

$$X_i(s) = \frac{1}{2} + \frac{1}{\pi} + \arcsin(\sin(\omega_i S))$$

Les choix des fréquences permet de s'assurer que tous les points de l'espace  $[0, 1]^p$  seront atteints, ce qui donne l'indépendance. Ce choix est détaillé dans la suite.

#### Décomposition de Fourier

Si les fréquences sont entières et positives alors nous avons

$$\mathbb{E}(Y) = \int_{[0,1]^p} f(x_1, \dots, x_p) d(x_1, \dots, x_p) = \frac{1}{2\pi} \int_{[-\pi, \pi]} f(x_1(s), \dots, x_p(s)) ds$$

Le changement de variables a donc permis de remplacer une intégrale multi-dimensionnelle par une intégrale simple.

Pour la variance nous avons le même changement de variables, ce qui permet d'écrire

$$\begin{aligned}
 \text{Var}(Y) &= \frac{1}{2\pi} \int_{[-\pi, \pi]} f^2(x_1(s), \dots, x_p(s)) ds - \mathbb{E}(Y)^2 \\
 &\stackrel{\text{Parseval}}{\simeq} \sum_{j=-\infty}^{+\infty} (A_j^2 + B_j^2) - (A_0^2 + B_0^2) \\
 &= 2 \sum_{j=1}^{+\infty} (A_j^2 + B_j^2) \tag{A.4.1}
 \end{aligned}$$

où  $A_j$  et  $B_j$  sont les coefficients de Fourier donnés par

$$\begin{aligned}
 A_j &= \frac{1}{2\pi} \int_{[-\pi, \pi]} f(x_1(s), \dots, x_p(s)) \cos(js) ds \\
 B_j &= \frac{1}{2\pi} \int_{[-\pi, \pi]} f(x_1(s), \dots, x_p(s)) \sin(js) ds
 \end{aligned}$$

### Indices d'ordre 1

Pour estimer la quantité  $\text{Var}(\mathbb{E}(Y/X_i))$  il suffit de sommer dans la formule (A.4.1) uniquement les termes correspondant aux harmoniques de la  $i^{\text{e}}$  fréquence, ce qui s'écrit

$$V_i = 2 \sum_{j=1}^{+\infty} (A_{j\omega_i}^2 + B_{j\omega_i}^2)$$

et l'indice d'ordre 1

$$S_i = \frac{V_i}{V}$$

est le même que celui rencontré avec la méthode de SOBOL.

### Indices totaux

SALTELLI propose de calculer les indices d'ordre totaux (méthode extended FAST) donnés par

$$S_i^{\text{tot}} = 1 - \frac{V_{\sim i}}{V}$$

où  $V_{\sim i}$  est la somme des harmoniques de toutes les fréquences autres que  $i^{\text{e}}$ . Le terme  $V_{\sim i}$  est donc égal à

$$V_{\sim i} = \sum_{j \neq i} V_j$$

Ce sont les mêmes indices que les indices totaux de SOBOL.

## A.5 Conclusion : dans la pratique

La démarche à suivre pour conduire une analyse de sensibilité est la suivante :

### 1. Modélisation de l'incertitude :

- déterminer la réponses de façon à ce qu'elle soit un bon indicateur du phénomène à étudier ; par exemple, si la sortie du code de calcul est une fonction du temps, une réponse possible est la valeur au temps final ;
  - déterminer les facteurs : faire la liste de toutes les sources d'incertitudes ; partir de la liste des paramètres du code de calcul est un bon début, mais il faut garder à l'esprit que l'incertitude ne concernera pas nécessairement les paramètres tels qu'ils apparaissent ; par exemple si les paramètres du code de calcul sont des coordonnées de points, l'incertitude concernera vraisemblablement la distance entre ces points ;
  - déterminer une distribution de probabilités pour chaque facteur ; pour cela faire appel à toutes les sources d'informations disponibles : littérature, estimation sur des données expérimentales, avis d'experts, ...
  - déterminer une structure de corrélation entre les facteurs.
2. **Analyse d'incertitude** : générer un échantillon, calculer la réponse correspondant et estimer certaines caractéristiques de la loi de la réponse (normalité, moments, ...) ; si la réponse varie peu il n'est pas nécessaire d'aller plus loin ;
  3. **Analyse de sensibilité** : choisir une méthode d'analyse en fonction des objectifs, générer le plan d'expériences correspondant cette méthode, calculer la réponse correspondant et estimer les indices de sensibilité.

Nous avons présenté quatre méthodes d'analyse de sensibilité :

1. **indices linéaires (PEAR, SRC, PCC)** : cette méthode permet de quantifier l'influence de chaque facteur ; ses avantages sont la simplicité et l'efficacité ; son inconvénient est qu'elle ne s'applique que si la régression linéaire est valide, ce qui limite son champ d'application ; l'adaptation à certains cas monotones avec la transformation sur les rangs étend ses possibilités ; remarquons que cette méthode ne nécessite pas de plan d'expériences particulier, il est donc possible de l'appliquer sur l'échantillon ayant servi à l'analyse d'incertitude ; c'est donc une méthode à tester avant d'essayer des méthodes plus lourdes ;
2. **Méthode de Morris** : cette méthode permet de classer les facteurs par ordre d'influence, de déterminer les facteurs non influents et d'identifier ceux ayant des effets non linéaires ou impliqués dans des interactions (sans distinction des deux cas) ; elle est très utile quand le modèle comporte beaucoup de facteurs car elle nécessite peu de simulations ; nous ne lui trouvons pas d'inconvénients et nous la recommandons avant toute étude approfondie ;
3. **Méthode de Sobol** : cette méthode permet de quantifier l'influence de chaque facteur (seul et influence totale), ainsi que les interactions de tous les ordres ; elle peut aussi

prendre en compte un modèle de corrélation avec les indices multidimensionnels ; c'est la méthode la plus complète ; son gros inconvénient est l'importante variabilité des estimations : pour avoir des estimations correctes il peut falloir 10 000 simulations par facteur ; nous conseillons donc d'utiliser cette méthode sur une surface de réponse ; cette dernière peut être construite à partir de l'échantillon de l'analyse d'incertitude ;

4. **Méthode FAST** : cette méthode permet d'estimer les mêmes indices que la méthode de SOBOL, mais elle nécessite beaucoup moins de simulations. C'est donc une alternative très intéressante.

L'ensemble de ces méthodes devrait permettre de conduire la plupart des analyses de sensibilité. Pour les autres méthodes, nous renvoyons le lecteur aux ouvrages de SALTELLI [SCS 00] (et particulièrement le chapitre 2 « Hitchhiker's Guide to Sensitivity Analysis »), ainsi que [STCR 04].



APPORT DES MÉTHODES PROBABILISTES DANS LA SIMULATION  
DU COMPORTEMENT SOUS IRRADIATION DU COMBUSTIBLE À PARTICULES

**Résumé :** Ce travail de thèse est consacré à l'évaluation d'espérances mathématiques et se situe plus particulièrement dans un contexte de fiabilité des structures. Nous cherchons à estimer une probabilité de défaillance (supposée faible) en tenant compte des incertitudes des paramètres influents du système. Notre objectif est d'atteindre un bon compromis entre la précision de l'estimation et le coût de calcul associé. L'application porte sur l'estimation de la probabilité de rupture du combustible à particules d'un réacteur nucléaire de type HTR, via un code de calcul numérique coûteux. Nous considérons différentes approches statistiques pour traiter le problème. Tout d'abord, nous examinons une méthode de Monte Carlo de réduction de variance : le tirage d'importance. Pour le cas paramétrique, nous proposons des algorithmes adaptatifs pour construire une suite de densités de probabilité qui convergerait vers la densité d'importance optimale. Nous présentons ensuite plusieurs estimateurs de l'espérance mathématique à partir de cette suite de densités. Par la suite, nous examinons une méthode de simulation multi-niveaux utilisant les algorithmes de Monte Carlo par Chaînes de Markov. Enfin, nous nous intéressons au problème connexe de l'estimation de quantile (non extrême) de quantité physique simulée par un code numérique coûteux. Nous proposons une méthode de stratification contrôlée où des réalisations des variables d'entrées sont échantillonnées dans des zones déterminées à partir d'un modèle réduit de la réponse. L'estimation du quantile se fait alors à partir de cet échantillon.

**Mots clés :** Fiabilité des structures, probabilité de défaillance, simulations de Monte Carlo, tirage d'importance, simulation multi-niveaux, estimation de quantile, stratification contrôlée, combustible à particules.

PROBABILISTIC METHODS FOR THE SIMULATION  
OF FUEL PARTICLES BEHAVIOR UNDER IRRADIATION

**Abstract :** This work is devoted to the evaluation of mathematical expectations in the context of structural reliability. We seek a failure probability estimate (that we assume low), taking into account the uncertainty of influential parameters of the system. Our goal is to reach a good compromise between the accuracy of the estimate and the associated computational cost. This approach is used to estimate the failure probability of fuel particles from a HTR-type nuclear reactor. This estimate is obtained by means of costly numerical simulations. We consider different statistical numerical methods to tackle the problem. First, we consider a (optimized) Monte Carlo method with a reducing variance technique : importance sampling. For the parametric case, we propose adaptive algorithms in order to build a series of probability densities that will eventually converge to optimal importance density. We then present several estimates of the mathematical expectation based on this series of densities. Next, we consider a multi-level method using Monte Carlo Markov Chain algorithm. Finally, we focus our attention to the related problem of quantile estimation (non extreme) of physical output from a large-scale numerical code. We propose a controlled stratification method. The random input parameters are sampled in specific regions obtained from surrogate of the response. The estimation of the quantile is then computed from this sample.

**Keywords :** Structural reliability, failure probability, Monte Carlo simulations, importance sampling, multi-level simulation, quantile estimation, fuel particles.